NASr-21(02)

MEMORANDUM
RM-5117-NASA
NOVEMBER 1966

MULTIPLE-ACCESS TECHNIQUES FOR COMMUNICATION SATELLITES: Analog Modulation, Frequency-Division Multiplexing, and Related Signal Processing Methods

E. E. Reinhart

PREPARED FOR:
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

The RAND Corporation
SANTA MONICA • CALIFORNIA

RAND RESEARCH BRIEF
THE RAND CORPORATION

December 1966

RB-5117

RM-5117-NASA, Multiple-Access Techniques for Communication Satellites: Analog Modulation, Frequency-Division Multiplexing, and Related Signal Processing Methods, E. E. Reinhart, RAND Memorandum, December 1966, 281 pp.

PURPOSE: To present basic theoretical data on possible methods of providing multichannel communication links between a satellite and an earth station.

RELATED TO: This Memorandum is one of a series designed to lay the groundwork for a comparison of methods by means of which a satellite-borne radio repeater can provide communication to several pairs of earth stations simultaneously. Alternative techniques were presented in RM-4997-NASA, Multiple-Access Techniques for Communication Satellites: Digital Modulation, Time-Division Multiplexing, and Related Signal Processing, September 1966. The requirements and constraints from the system designer's viewpoint were presented in RM-4298-NASA, Multiple-Access Techniques for Communication Satellites: I. Survey of the Problem.

MODULATION/DEMODULATION METHODS: This Memorandum unifies, refines, and extends work hitherto scattered in engineering journals. The concepts of modern communications theory are introduced as needed to provide a systematic tutorial development of the performance characteristics of six analog modulation methods, including the single sideband form of FM. A unified treatment is achieved by regarding each modulation method as a special case of a general carrier that has been modulated in both amplitude and frequency. The analytic signal representation is used to describe both the modulated carrier and the output of the basic analog demodulators. For each modulation/demodulation method, expressions giving the dependence of RF power and bandwidth on the output message quality requirement are derived for realistically arbitrary messages and modulation parameters. These are tabulated both for single channel circuits and for SSB frequency-division multiplexed messages. Special attention is given to the potential of FM receivers employing frequency feedback loops for extending the range over which bandwidth can be traded for power.

SIGNAL PROCESSING: Preemphasis and companding improve transmission efficiency by modifying the signal at the transmitter and then restoring it to the detriment of the noise at the receiver. The maximum improvement possible by preemphasizing the baseband depends on the spectrum of the signal and of the noise. For single speech channels, optimum preemphasis can save up to 3 db with AM and up to 10 db with FM. For multichannel speech basebands, conventional preemphasis yields no saving with AM and only about 2 db with FM. This could be increased by 3 db for each by appropriately preemphasizing each channel individually before multiplexing.

Syllabic companding can be applied with any modulation method. Its quieting effect between syllables and words can be used on multichannel links to produce greater bandwidth or power savings than does preemphasis.

TASI, an elaborate switching system, doubles the capacity of undersea telephone cables by switching channels during gaps in speech. In future satellite links, TASI could also be employed to reduce the RF power or bandwidth needed to transmit a given number of voice channels.

MW/EER

NASr-21(02)

MEMORANDUM
RM-5117-NASA
NOVEMBER 1966

MULTIPLE-ACCESS TECHNIQUES FOR COMMUNICATION SATELLITES: Analog Modulation, Frequency-Division Multiplexing, and Related Signal Processing Methods

E. E. Reinhart

This research is sponsored by the National Aeronautics and Space Administration under Contract No. NASr-21. This report does not necessarily represent the views of the National Aeronautics and Space Administration.

The RAND Corporation
1700 MAIN ST. • SANTA MONICA • CALIFORNIA 90406

PREFACE

This Memorandum continues the RAND study of multiple-access techniques for communication satellites undertaken for the National Aeronautics and Space Administration. As background for future comparisons of multiple-access methods, it provides a tutorial treatment of the theory of analog modulation, frequency-division multiplexing, and the signal processing techniques of preemphasis, companding, and time assignment speech interpolation. The principal objective is to derive general expressions for the RF bandwidth and the peak and average carrier power required to transmit arbitrary numbers and types of messages with given fidelity over a single radio link. In these derivations, only additive gaussian noise and above-threshold signals are considered; the effects of signal distortion, cross-talk, and interference of various kinds are not treated. The results can be used to compare the performance of the various analog techniques both with each other and with the digital techniques discussed in a companion Memorandum, RM-4997-NASA.[2]

SUMMARY

This Memorandum provides an elementary, comparison-oriented description of analog modulation methods and the techniques of multiplexing and signal processing normally employed with them.

The various forms of amplitude and angle modulation (AM, DSB, SSB, PM, FM, SSBFM) are defined as special cases of a general carrier that has been modulated simultaneously in both amplitude and phase or frequency. These definitions establish for each modulation method the relationship between the message (or baseband) waveform and the resultant RF waveform.

The corresponding relationships between Fourier spectra are then developed with the aid of the analytic signal representation of the modulated carrier. In this representation, the modulation method is characterized by the functional dependence on the baseband of a single complex waveform called the modulation function whose spectrum is closely related to that of the modulated carrier.

To describe the modulated carrier in the typical case where the baseband signal carries information, the concepts of probability density function and power spectrum are introduced and used to express the RF peak-to-average power ratio and bandwidth in terms of the corresponding baseband quantities. These expressions are illustrated through application to sinusoidal and gaussian distributed basebands.

After discussing the statistical characteristics of noise, general relationships between input and output signal-to-noise ratios are derived for superheterodyne receivers employing each of the four

basic types of demodulators: product, envelope, phase, and frequency. The effectiveness of frequency feedback for extending the threshold of FM receivers is also reviewed.

Combining the signal-to-noise relationships with those for peak power and bandwidth, expressions are obtained for comparing analog modulation methods in terms of their cost in RF power and bandwidth for any specified baseband signal and output signal-to-noise ratio. These expressions are tabulated for convenient reference and may be used directly to compare the relative efficiency of the analog modulation methods for single channel transmission.

Similar signal-to-noise expressions are also developed for multichannel basebands formed by the most commonly used type of frequency-division multiplexing, namely that which employs single-sideband modulated subcarriers to represent the individual channel inputs. Particular attention is devoted to the statistical properties of such basebands in the important practical case where the inputs are commercial-quality telephone channels.

Finally, the Memorandum examines the extent to which the performance of frequency-division multiplexed, analog modulated radio links can be improved through application of such signal processing techniques as preemphasis, companding, and TASI (time assignment speech interpolation). In each case, methods are developed for estimating the reduction in RF power and/or bandwidth made possible by the technique.

ACKNOWLEDGMENTS

The author acknowledges with gratitude many valuable discussions with E. Bedrosian and C. R. Lindholm of the RAND Corporation.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

Most of the subscripted and superscripted symbols in this Memorandum are designed to be self-explanatory: the symbol identifies the physical nature of the quantity, the subscript(s) tell the point at which it is measured, and the superscript(s) (if any) identify the type(s) of signal processing associated with the quantity. The letters used in this way are defined in the list on this and the following page.

Symbol	Quantity
A	unmodulated carrier amplitude
B	signal bandwidth
E(t)	signal waveform (voltage or current)
e(t)	normalized signal waveform
K	signal-to-noise power ratio
N	average noise power
S	average signal power
$\hat{S}$	peak signal power
U(t)	noise waveform
u(t)	normalized noise waveform
V	volume, equivalent volume
X(t)	signal-plus-noise waveform
x(t)	normalized signal-plus-noise waveform
Λ	peak-to-average power ratio
φ	initial phase

Subscript	Point of Measurement
B	baseband at modulator input
C	single channel output of demultiplexer
I	IF amplifier output

L	low-pass filter output of demodulator
M	modulator output
R	receiver input
S	single channel input to multiplexer
T	transmitter output
Xi	i^{th} single channel at point X
Xo	unmodulated carrier at point X
o	oscillator input to modulator

Superscript	Type of Signal Processing
C	compression
D	deemphasis
E	expansion
P	preemphasis

Complete Alphabetical List

A_o	amplitude of unmodulated carrier
$A(t)$	instantaneous amplitude of modulated carrier
A_X	amplitude of unmodulated carrier at point X
a	amplitude
$a(t)$	amplitude ratio of modulated carrier
B	subscript indicating baseband output of multiplexer
B_{Nc}, B_{No}	closed-loop noise bandwidth, open-loop noise bandwidth
B_X	signal bandwidth at point X
$b(t)$	baseband function
C	subscript indicating single-channel output of demultiplexer
C	superscript indicating compression
C, C^P	proportionality factor
$C_{x_1 x_2 \ldots}(\xi_1, \xi_2 \ldots)$	characteristic function of random variables $x_1, x_2, \ldots$
$c(t)$	complex modulated carrier
D	superscript indicating deemphasis
D	deviation ratio or modulation index for FM and SSBFM
D_a	maximum value of deviation ratio due to spectral allocation
D_t	threshold value of deviation ratio
d	constant
E	superscript indicating expansion
$E_X(t)$	signal voltage or current at point X
$E_X(\omega), E_X(f)$	Fourier transform of $e_X(t)$
$E_X^Y(t)$	signal waveform at point X after signal processing identified by Y
$e(t)$	normalized modulated carrier voltage or current
$e_X(t)$	normalized signal voltage or current at point X

F	feedback factor
f	frequency
$f( )$	arbitrary function
f_d	rms frequency deviation
f_i	subcarrier frequency in i^{th} channel
f_ℓ	minimum baseband frequency
f_m	maximum baseband frequency
f_o	carrier frequency
f_p	frequency at which power spectrum has its maximum value
G_I	power gain of IF amplifier
G_T, G_R	gains of transmitting, receiving antennas
$g( )$	arbitrary function
g_P	power spectral response of preemphasis filter
$H_c(f)$	closed-loop transfer function of feedback FM receiver
$H_D(f)$	transfer function of deemphasis network
$H_F(f)$	transfer function of feedback amplifier in feedback FM receiver
$H_I(f)$	transfer function of IF amplifier
$H_{IL}(f)$	transfer function of low pass equivalent to IF amplifier
$H_o(f)$	open-loop transfer function of feedback FM receiver
$H_P(f)$	transfer function of preemphasis network
I	subscript indicating IF amplifier output
I_P	preemphasis improvement factor
$I_{P\ell}$	minimum or optimum value of I_P
i	subscript identifying i^{th} single channel
j	imaginary unit
K_X	ratio of average signal power to average noise power in the signal bandwidth at point X
K_X^Y	signal-to-noise ratio at point X after signal processing identified by Y

k	Boltzmann's constant
k_1, k_2	proportionality constants
L	path or propagation loss
ℓ	subscript indicating lowest or minimum value
M	modulation index for AM, DSB, SSB
M_n	n^{th} Fourier coefficient of modulation function m(t)
$M(\omega)$	Fourier transform of m(t)
m	subscript indicating maximum value
$m(t)$	modulation function
N_o	noise power density at receiver input
N_X	average noise power measured at point X
N_X^Y	average noise power at point X in presence of signal processing Y
N_{Xi}	average noise power in i^{th} channel at point X
o	subscript indicating unmodulated carrier
P	superscript indicating preemphasis
$p(t)$	in-phase component of normalized noise waveform
$p_X(x)$	probability density function of random variable X
Q	total number of channels in baseband
q	number of active channels in baseband
$q(t)$	quadrature component of normalized noise waveform
$R_x(\tau)$	autocorrelation function of random process x(t)
$R_{xy}(\tau)$	cross correlation function of random processes x(t) and y(t)
$r(t)$	envelope of normalized signal-plus-noise waveform
S_X	average signal power measured at point X
$S_1(Q)$	power level exceeded by average baseband power one percent of time

S_X^Y	average signal power at point X with signal processing Y
S_{Xo}	average signal power of unmodulated carrier at point X
S_{Xi}	average signal power in i^{th} channel at point X
T	period
T_s	system noise temperature
t	time
t	subscript indicating threshold
$U_X(t)$	noise waveform at point X
u	integration variable
$u_X(t)$	normalized noise waveform at point X
V_X	volume or equivalent volume at point X
v(t)	envelope of normalized noise waveform
$W_X(f)$	power spectrum of random process X(t)
w(f)	normalized power spectrum
w_N	normalized noise power spectrum at demodulator output
w_S	normalized signal power spectrum at demodulator output
X_n	n^{th} Fourier coefficient of waveform x(t)
$X(\omega), X(f)$	Fourier spectrum or transform of x(t)
$X_X(t)$	signal-plus-noise waveform at point X
x	variable of integration, argument of probability density function
x(t)	arbitrary waveform
$x_X(t)$	normalized signal-plus-noise waveform at point X
y(t)	arbitrary waveform
Z_n	n^{th} Fourier coefficient of waveform z(t)
$Z(\omega)$	Fourier spectrum or transform of z(t)
z(t)	arbitrary complex waveform

$\mathfrak{A}$	overload probability in definition of peak amplitude
β	probability used in defining (peak) bandwidth
Γ_i	baseband load factor referred to i^{th} channel
γ	modulation factor of VCO in a feedback FM receiver
γ_i	factor of proportionality between average carrier power and the SNR at the output of the i^{th} channel
δ	slope factor of discriminator in a feedback FM receiver
$\delta(f)$	Dirac delta function
ϵ	overload probability in definition of peak envelope
$\epsilon(t)$	ratio of noise waveform to signal-dependent waveform at envelope demodulator output
$\eta(t)$	incremental phase deviation at demodulator output due to noise
$\Theta(f)$	Fourier transform of $\theta(t)$
$\theta(t)$	phase angle of arbitrary complex waveform phase angle of signal-plus-noise waveform
$\Lambda_B(Q)$	the 0.1 percent peak-to-average baseband power ratio for Q-channel baseband
Λ_X	peak-to-average power ratio at point X
ν	normalized baseband frequency
ξ	argument of characteristic function
ρ	ratio of baseband SNR before deemphasis to its value after deemphasis
ρ_ℓ	smallest value of ρ
$\rho(t)$	incremental amplitude modulation at envelope demodulator output due to noise
σ_X	standard deviation of random variable X
τ	delay variable in autocorrelation function, activity factor
Φ	modulation index for PM
$\Phi_o(t)$	phase of unmodulated carrier

$\Phi(t)$	instantaneous phase of modulated carrier
$\dot{\Phi}(t)$	instantaneous angular frequency deviation
φ	phase constant
φ_o	initial phase of unmodulated carrier
$\varphi(t)$	phase deviation of modulated carrier
$\dot{\varphi}(t)$	angular frequency deviation of modulated carrier
$\psi(t)$	phase deviation of noise waveform
Ω	peak angular frequency deviation
ω	angular velocity, angular frequency
ω_o	carrier angular frequency

Mathematical Functions and Operators

arg[z(t)]	argument of z(t)
erfc( )	complementary error function
exp( )	exponential function
$\mathfrak{F}$	Fourier transform
$\mathscr{H}$	Hilbert transform
$I_o( )$	Bessel function of order zero with imaginary argument
$J_n( )$	Bessel function of order n
ln	natural logarithm
log, lg	common logarithm
Pr { }	probability of the event described in the brackets { }
sgn( )	signum function
$\hat{x}$	peak value of x(t)
$\underline{x}(t)$	analytic signal representation of x(t)
$\overline{x(t)}$	average or mean value of x(t)
$\check{x}(t)$	Hilbert transform of x(t)
$\dot{x}(t)$	time derivative of x(t)

$x(t)*y(t)$	convolution of $x(t)$ and $y(t)$
$z^*(t)$	complex conjugate of $z(t)$
$\|z(t)\|$	absolute value of $z(t)$

Abbreviations

AM	amplitude modulation
DSB	double-sideband suppressed carrier amplitude modulation
FDM	frequency-division multiplexing
FM	frequency modulation
IF	intermediate frequency
LSB	lower sideband
PEP	peak envelope power
PM	phase modulation
RF	radio frequency
SNR	signal-to-noise ratio
SSB	single-sideband suppressed carrier amplitude modulation
SSBFM	single-sideband frequency modulation
SSBPM	single-sideband phase modulation
USB	upper sideband
VCO	voltage controlled oscillator

I. INTRODUCTION

OBJECTIVES AND SCOPE

When a satellite-borne radio repeater can be used to complete a communication circuit between two earth stations, the stations are said to have "access" to the repeater. When more than two earth stations can have independent access to the same repeater simultaneously, the satellite is said to provide multiple access. Even when in a low-altitude orbit, a satellite-borne repeater is potentially accessible to all suitably equipped earth stations within a relatively large geographic area. It is obvious that techniques which enable repeaters to realize their multiple-access potential efficiently are essential to the full exploitation of satellites in future communication systems.

This Memorandum is one in a series whose ultimate goal is the description and comparison of a number of alternative multiple-access techniques for typical user needs. The first of this series[1] was concerned primarily with the factors which define the multiple-access problem from the communication system designer's point of view, including the geographical distribution of potential users, expected traffic demands, orbit considerations, booster capabilities, the state-of-the-art in amplifier design, frequency allocations, and atmospheric propagation effects. Some of these factors indicate the types and numbers of communication links that future systems should provide, others impose restrictions on the technical design of such systems, but all are largely beyond the control of the system designer.

The present Memorandum is intended to provide basic information for some of the design choices which <u>are</u> within the purview of the

system engineer in devising appropriate multiple-access techniques. The treatment is tutorial in character and is addressed to those responsible for overall system planning rather than to the specialist in modulation theory. Although many of the concepts and analytic tools of modern communication theory are employed, no previous familiarity with them is presumed.

The principal objective is to provide a self-contained and unified introduction to the theory of several of the basic techniques of modulation, multiplexing, and signal processing, with results expressed in terms which facilitate comparisons of the techniques when combined in various ways to provide multichannel links between a satellite and an earth station. For this purpose, a systematic development is given for the signal-to-noise relationships needed to determine how the quality of an arbitrary message at the output of any single channel in such a link depends on the characteristics of all the messages, on the capacity of the link, and on the particular technical choices embodied in the design of the link terminals. The performance of a given combination of the basic techniques may then be measured by the costs--in RF bandwidth and power (both peak and average)--of achieving a specified single channel message quality.

A few restrictions on the scope of the treatment should also be noted. Only analog modulation methods, frequency-division multiplexing, and the signal processing techniques appropriate to them are considered.* In developing the signal-to-noise relationships needed

---

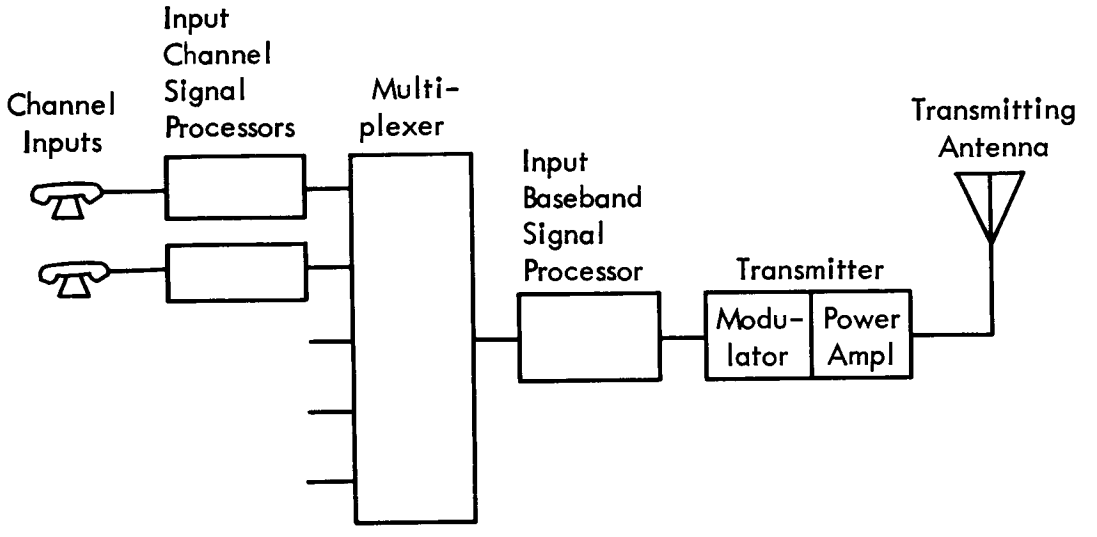
*Digital modulation, time-division multiplexing, and related signal processing methods are the subject of another Memorandum in this series.(2)

to compare these methods the signal is assumed to be much larger than the noise, and idealized circuits such as distortionless filters with rectangular passbands, perfectly linear amplifiers, and exactly synchronous demodulators are usually assumed. Link performance comparisons based on these equations will thus represent theoretical goals which can only be approached in practice. There is little if any discussion of how the assumed ideal circuits can be approximated in practice, and the comparisons do not take into account important practical considerations such as the weight, size, reliability, and dollar cost implications of alternative design choices. Finally, no attempt is made in this Memorandum to describe and compare the performance of multiple-access schemes per se. This will be treated in subsequent reports in this series.

ORGANIZATION OF MATERIAL

The techniques to be examined in this Memorandum are divided into three categories: modulation, multiplexing, and signal processing. The distinction between them can be illustrated by reference to the functional block diagram shown in Fig. 1. It represents a one-way, multichannel, radio communication link, and emphasizes those link elements in which significant signal transformations (as opposed to simple amplification or attenuation) take place.

Under the heading of "modulation" is included the complementary pair of signal transformations that occur in the blocks labeled modulator and demodulator--respectively the translation to radio frequencies of a representation of the composite multichannel signal or baseband at the transmitter and its subsequent recovery in the



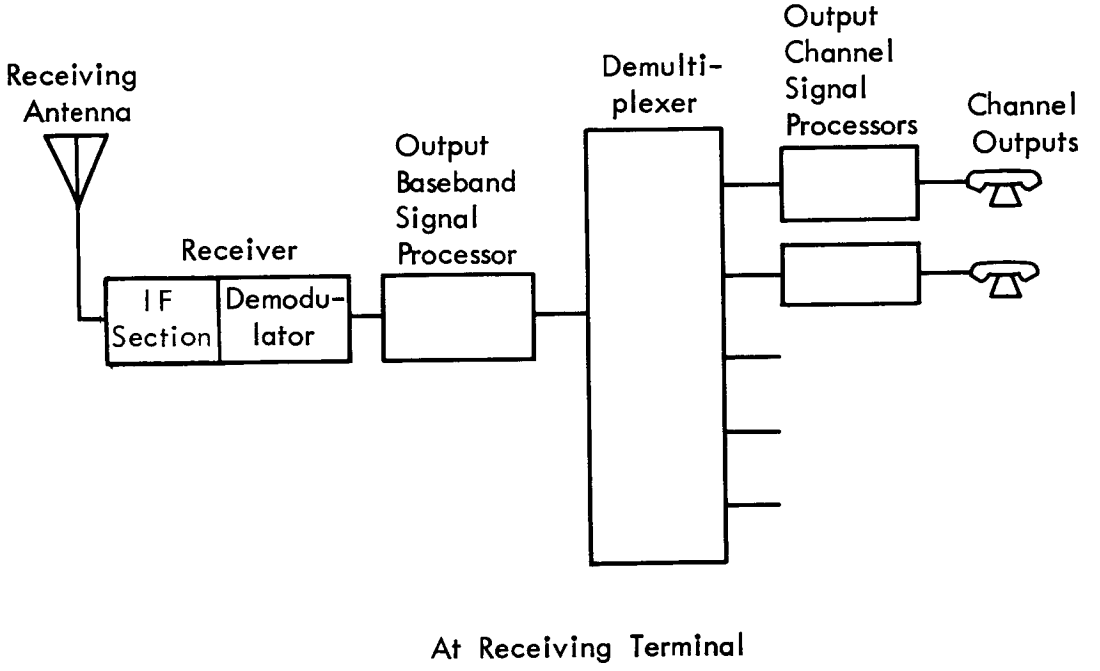


Fig. 1—Generalized block diagram of a radio communication link (for one direction of transmission)

receiver. Likewise, the multiplexing category includes not only the operation of combining a number of single channel input signals to form a baseband signal in the block labeled multiplexer, but also the inverse operation of resolving the baseband into its component channels in the demultiplexer. Finally, "signal processing" embraces all of the other signal transformations at the transmitter and receiver whether they are applied to the individual channels or to the baseband. A separate section of this Memorandum is devoted to each of the three categories.

In Section II, the properties of several amplitude and angle modulation and demodulation methods are derived, with emphasis on the relationships among the parameters which characterize the modulating baseband, the modulated RF signal, the circuits of the modulator and demodulator, and, most important, the quality of the signal at the demodulator output as measured by the signal-to-noise ratio (SNR) at this point.

The discussion of modulation techniques precedes that for multiplexing and signal processing for two reasons. Firstly, a knowledge of modulation and demodulation methods is all that is required for predicting the performance of the simplest circuits of practical importance: those that carry only a single unprocessed message channel. Secondly, a knowledge of modulation techniques forms a necessary prerequisite to an analysis of most multiplexing methods and provides the motivation for certain types of signal processing.

The specific modulation techniques considered are: conventional amplitude modulation (AM), double-sideband suppressed carrier

amplitude modulation (DSB), single-sideband suppressed carrier amplitude modulation (SSB), phase modulation (PM), frequency modulation (FM), and the recently discovered single-sideband form of FM (SSBFM).[3,4] For each of these methods, the relations between the baseband and the RF signal characteristics (bandwidth and peak-to-average power ratio) are illustrated for two basebands of practical interest: a sinusoidal test tone and a noise-like signal which typifies the result of multiplexing a large number of independent channels whatever their individual nature. The expressions for output SNR developed for the four basic types of demodulators (product, envelope, phase, and frequency) may then be used directly to compare the performance of the different modulation methods when used for single channel circuits.

Section III, on multiplexing, begins with a discussion of frequency-division multiplexing (FDM), and in particular, the type of FDM that employs SSB to place the component channels in their respective frequency slots. Using the relationships developed for SSB in Section II to represent the multiplexing and demultiplexing operations, separate expressions are derived for the single channel output SNRs for links which employ, for RF transmission, each of the modulation techniques discussed in Section II.

These SNR expressions involve not only parameters descriptive of the modulator and demodulator but also those which relate the single channel messages to the baseband. The latter parameters are investigated for the important special case where the single channel inputs carry human speech. This analysis of multichannel speech basebands takes explicit account of the fact that the individual speech channels

are busy* only a part of the time, and when busy, do not carry speech power either continuously or at the same average power level. The material in Section III thus provides a basis for comparing the performance of the various modulation methods when used for multichannel voice transmission.

Section IV is concerned with the types of signal processing that are sometimes employed to improve transmission efficiency on analog communication circuits. The techniques considered in this section include preemphasis, companding, and TASI (time assignment speech interpolation). The object of the discussion in each case is to describe the changes in signal characteristics produced by processing and to provide quantitative relationships for evaluating the resultant improvement in link efficiency when used with various modulation methods. These relations are also illustrated for the special case of multichannel voice basebands.

In a report such as this, where expressions are developed for different types of signals, at several different points in several kinds of systems, a consistent and unambiguous system of notation is particularly important. This Memorandum uses subscripted and superscripted symbols in which the symbol identifies the physical nature of the quantity, the subscript tells the point at which it is measured, and the superscript (if any) describes the signal processing applied to the quantity. The letters used in this way are defined at the beginning of the List of Symbols.

---

*A channel is busy when it is unavailable for use by other subscribers.

II. ANALOG MODULATION METHODS

INTRODUCTION

In communication engineering, the term "modulation" normally refers to the process in which an information-bearing signal called the modulating signal, message, or baseband* is combined with a higher frequency signal called the carrier to form a new signal called the modulated carrier whose properties are better suited to transmission over the medium between transmitter and receiver. To be of practical interest, the modulation process must be such that the baseband can subsequently be recovered from the modulated carrier without objectionable distortion. The modulation process is said to be analog if the baseband signal is allowed to vary over a continuous range of values, and digital if it is restricted to a discrete set of values.

The process is called modulation because classically it has been viewed as one in which some property of the carrier, such as its amplitude or frequency, was modulated by, i.e., made to vary directly with, the baseband signal. Recently, however, there has been an increasing tendency to regard modulation as the process of operating directly on the baseband,[4] rather than on the carrier, to obtain a new signal whose spectrum is then translated unchanged to the carrier frequency. This latter view of modulation is based on the fact that all analog modulated carriers can be represented mathematically as the product of a transformed baseband signal and the unmodulated carrier.

---

*The term "baseband" usually connotes a composite modulating signal, obtained by multiplexing two or more single channel inputs. In this report, however, the modulating signal will be referred to as the baseband even when it includes only a single channel.

It has been particularly fruitful in showing how, in the older view of modulation, the carrier amplitude and phase must be varied to yield modulated carriers with new and potentially useful characteristics. Moreover, it emphasizes the primary role of modulation: to translate the information-bearing signal to a more desirable place in the spectrum, whether it be to a particular frequency slot in a frequency-division multiplexed baseband or to a radio frequency channel for subsequent transmission over a cable or through the atmosphere.

In this section, both views of the modulation process are employed to describe the signal that results when an arbitrary baseband signal analog modulates a sinusoidal carrier in various ways. First, the classical approach is used to define the different modulation methods in terms of the functional dependence of the modulated carrier waveform on the baseband waveform. Then the more recent signal-product concept is introduced to provide a basis for discussing the spectral characteristics of modulated carriers and to show the origin of the simultaneous amplitude and phase modulations used to define the single-sideband methods.

Next, the modulated RF carrier is described in statistical terms for the normal situation in which the baseband itself can only be described statistically. Expressions are derived for the peak-to-average power ratio and the bandwidth of the RF signal in terms of the corresponding baseband quantities. The results are illustrated by applying them to both sinusoidal and gaussian distributed baseband signals. The sinusoid is commonly used for system alignment and testing and the noise-like signal is a good approximation to many multichannel basebands.

After describing the signal that results from various types of modulation, it is shown how the inverse operation of demodulation affects the combination of modulated carrier and noise which reaches the demodulator input terminals. Towards this end, simple mathematical representations of noise and of various types of demodulators are introduced and used to derive expressions for the ratio of average signal power to average noise power at the demodulator output. These expressions show how, for each modulation-demodulation method, this measure of output signal quality depends on the nature of the baseband signal, the modulation method, the receiver sensitivity, and the average signal power that reaches the receiver input terminals. The relations between average input signal power and output SNR are also combined with the equations for peak-to-average carrier power to show how the output signal quality depends on the peak received power for each modulation method.

MODULATION AND THE MODULATED CARRIER

Modulation as an Operation on the Carrier--Waveform Relations

Adopting the classical view for the moment, modulation is the operation at the transmitting terminal (see Fig. 1) in which the baseband signal $E_B(t)$ is impressed on one or more of the parameters of a carrier $E_o(t)$. In this report, the carrier waveform is assumed to be sinusoidal with amplitude A_o, frequency f_o, and initial (zero time) phase φ_o:

$$E_o(t) = A_o \cos \Phi_o(t) \tag{1}$$

where

$$\Phi_o(t) = \omega_o t + \varphi_o$$

is the total phase angle, or simply the "phase" of the unmodulated carrier. The time derivative of phase is the angular frequency:

$$\dot{\Phi}_o(t) = \omega_o = 2\pi f_o$$

Regardless of the nature of the modulation, the modulated carrier waveform $E_M(t)$ that appears at the modulator output can always be written in an analogous form

$$E_M(t) = A(t) \cos \Phi(t)$$

where $A(t)$ is called the instantaneous amplitude and $\Phi(t)$ the instantaneous phase of the modulated carrier. The magnitude $|A(t)|$ of the instantaneous amplitude is called the carrier envelope because it bounds the excursions of the modulated carrier. The time derivative $\dot{\Phi}(t)$ of the instantaneous phase is called the instantaneous angular frequency because, just as $\omega_o/2\pi = f_o$ gives the number of complete oscillations per second executed by the unmodulated carrier, $\dot{\Phi}(t)/2\pi$ gives the time-dependent oscillation rate of the modulated carrier at time t.

The effect of modulation on the carrier is thus to change its amplitude from A_o to $A(t)$, its phase from $\Phi_o(t)$ to $\Phi(t)$, and its angular frequency from ω_o to $\dot{\Phi}(t)$. Using the unmodulated carrier for reference, these changes are respectively measured by the following quantities:

$$\left.\begin{array}{ll} \text{amplitude ratio} & a(t) \equiv A(t)/A_o \\ \text{phase deviation} & \varphi(t) \equiv \Phi(t) - \Phi_o(t) \\ \text{angular frequency deviation} & \dot{\varphi}(t) \equiv \dot{\Phi}(t) - \omega_o \end{array}\right\} \qquad (2)$$

In terms of the amplitude ratio and phase deviation, the modulator output is then

$$E_M(t) = a(t)\, A_o \cos\left[\omega_o t + \varphi_o + \varphi(t)\right] \qquad (3)$$

Thus, if amplitude and phase are regarded as the carrier parameters to be modulated, the various analog modulation methods can be defined by specifying how $a(t)$ and $\varphi(t)$ depend on the baseband $E_B(t)$. The modulated carrier is then completely described whenever the baseband and unmodulated carrier are known exactly. Alternatively, if amplitude and angular frequency are adopted as the basic modulation parameters, it is the dependence of $a(t)$ and $\dot{\varphi}(t)$ on $E_B(t)$ that defines a modulation method. In this case, exact knowledge of $E_B(t)$ and $E_o(t)$ determines the modulated carrier waveform to within an additive phase constant (which may be neglected without loss of generality).

The two alternatives are essentially equivalent; in Table 1, both are used to define the modulation methods to be considered in this study. Before discussing these definitions, however, it should be recognized that a knowledge of the modulation parameters $a(t)$ and $\varphi(t)$ (or $a(t)$ and $\dot{\varphi}(t)$) determines the modulated carrier waveform not only at the modulator output but also at all subsequent points in the link up to and including the receiver input.*

This conclusion stems from the assumption that all link elements between the modulator and the receiver (power amplifier, transmission lines, antennas and propagation medium) are distortionless and time invariant. With this assumption, $a(t)$ gives at any point the ratio of the instantaneous amplitude of the modulated carrier to the

*It will subsequently be shown that $a(t)$ and $\varphi(t)$ also define the signal component of the waveform at the demodulator input.

Table 1

DEFINITION OF ANALOG MODULATION METHODS AS OPERATIONS ON THE CARRIER PARAMETERS

Analog Modulation Method	Amplitude Ratio $a(t) \equiv A(t)/A_o$	Phase Deviation $\varphi(t) \equiv \Phi(t) - \Phi_o(t)$	Angular Frequency Deviation $\dot{\varphi}(t) \equiv \dot{\Phi}(t) - \omega_o$	Normalized Modulated Carrier $e(t) = a(t)\cos[\omega_o t + \varphi(t)]$
AM	$1 + M\, e_B(t)$	0	0	$[1 + M\, e_B(t)]\cos \omega_o t$
DSB	$M\, e_B(t)$	0	0	$M\, e_B(t)\cos \omega_o t$
SSB[a]	$M\sqrt{e_B^2(t) + \breve{e}_B^2(t)}$	$\pm\tan^{-1}\dfrac{\breve{e}_B(t)}{e_B(t)}$	$\pm\dfrac{e_B(t)\dot{\breve{e}}_B(t) - \dot{e}_B(t)\breve{e}_B(t)}{e_B^2(t) + \breve{e}_B^2(t)}$	$M\sqrt{e_B^2(t)+\breve{e}_B^2(t)}\cos[\omega_o t \pm \tan^{-1}\dfrac{\breve{e}_B(t)}{e_B(t)}]$
PM	1	$\Phi\, e_B(t)$	$\Phi\, \dot{e}_B(t)$	$\cos[\omega_o t + \Phi\, e_B(t)]$
FM	1	$\Omega \int e_B(t)dt$	$\Omega\, e_B(t)$	$\cos[\omega_o t + \Omega \int e_B(t)dt]$
SSBFM[a]	$e^{\mp\Omega \int \breve{e}_B(t)dt}$	$\Omega \int e_B(t)dt$	$\pm\, \Omega\, e_B(t)$	$e^{\mp\Omega \int \breve{e}_B(t)dt}\cos[\omega_o t + \Omega\int e_B(t)dt]$

[a]Where alternative signs (±) appear, the upper sign yields a modulated carrier with an upper sideband; the lower sign a modulated carrier with a lower sideband.

amplitude that would exist there if the unmodulated carrier $E_o(t)$ had been transmitted. Similarly, $\varphi(t)$ gives the difference between the instantaneous phase of the modulated carrier and the phase that would exist at the same point if $E_o(t)$ had been transmitted.

For example, if at the transmitter output terminals, the amplitude and phase of the unmodulated carrier are respectively A_T and $\omega_o t + \varphi_T$ then the modulated carrier at this point is

$$E_T(t) = a(t)\ A_T \cos\left[\omega_o t + \varphi_T + \varphi(t)\right] \tag{4}$$

Likewise, if upon arrival at the receiver input terminals, the amplitude and phase of the unmodulated carrier are respectively A_R and $\omega_o t + \varphi_R$ then the signal input to the receiver is

$$E_R(t) = a(t)\ A_R \cos\left[\omega_o t + \varphi_R + \varphi(t)\right] \tag{5}$$

Comparing Eqs. (3), (4), and (5), it follows that, when normalized to the local value of the unmodulated carrier amplitude, the modulated carrier can be represented to within an additive phase constant by the same waveform at all points, viz:

$$e(t) = a(t) \cos\left[\omega_o t + \varphi(t)\right] \tag{6}$$

Explicit expressions for e(t) in terms of the baseband are included in Table 1 to permit a direct comparison of the modulated carriers resulting from the different modulation methods.

Once the normalized waveform e(t) is known, the modulated carrier at a given point is easily obtained through multiplication by the unmodulated carrier amplitude at that point. For this purpose it is useful to note that the amplitude at one point in a link may be found from that at another if the net power gain between the two points is known. For example, in the case of the transmitter output and receiver input just considered, the relation between unmodulated carrier amplitudes is

$$A_R = A_T \sqrt{G_T G_R / L}$$

where G_T and G_R are the gains of the transmitting and receiving antennas (including line losses) and L is the total propagation loss in the intervening medium.

Returning now to Table 1, two related waveforms require explanation. The first, denoted by $e_B(t)$ and appearing in nearly every entry, represents the baseband waveform $E_B(t)$ after normalization to its peak amplitude $\hat{E}_B$:

$$e_B(t) \equiv E_B(t) / \hat{E}_B$$

Here, $\hat{E}_B$ is defined as the value which $|E_B(t)|$ exceeds with some specified small probability. Since the baseband is assumed to have no dc component $\left(\overline{E_B(t)} = 0\right)$, one effect of normalization is to make the modulator constants of proportionality M, Φ, and Ω respectively equal to the peak deviations from the mean of $a(t)$, $\varphi(t)$, and $\dot{\varphi}(t)$. Another effect is to make the mean square value of the normalized

baseband signal equal to the ratio of average to peak baseband power

$$\overline{e_B^2(t)} = S_B/\hat{S}_B \equiv \Lambda_B^{-1}$$

The other waveform, which appears only in the entries for the single-sideband methods SSB and SSBFM, is denoted by $\breve{e}_B(t)$. It is called the Hilbert transform[5]* or harmonic conjugate of the normalized baseband $e_B(t)$. In general, for an arbitrary waveform $x(t)$ with finite total energy $\int_{-\infty}^{\infty} x^2(t)\,dt$, the Hilbert transform may be obtained by applying to $x(t)$ an integral operator $\mathcal{H}$ defined as the convolution of $x(t)$ with $(\pi t)^{-1}$:

$$\breve{x}(t) = \mathcal{H}[x(t)] \equiv \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{x(u)}{t-u}\,du \qquad (7)$$

In this definition, the Cauchy principal value of the integral is to be taken due to the singularity of the integrand at the origin. A delay-line analog of the integral may be constructed to derive $\breve{x}(t)$ from $x(t)$ when it is needed in practical circuits.** The appearance of the Hilbert transform in the expressions for the single-sideband methods is closely related to its spectral properties which will be considered later in this section in the discussion of modulated carrier spectra.

---

*The notation used here is that of Papoulis.[6] The more common notation, $\hat{x}(t)$, has been preempted in this report to identify "peak value."

**As will be shown later, $x(t)$ may also be derived from $x(t)$ using a 90° phase shifter.

Referring now to the specific entries in Table 1, several similarities and distinctions between modulation methods should be noted. To begin with, AM and DSB are very much alike. In each case, the amplitude ratio a(t) is a linear function of the modulating baseband and there is no phase or frequency deviation. Both AM and DSB may thus be described as pure amplitude modulation techniques. The only difference between them is that in AM, it is the deviation $A(t) - A_o$ of the instantaneous carrier amplitude from its unmodulated value that is proportional to the baseband, whereas in DSB it is A(t) itself.

In each case, the degree or depth of modulation is specified by the constant of proportionality M, which is called the modulation index. For AM, M is seen to be the ratio of the peak instantaneous amplitude deviation to the peak unmodulated carrier amplitude

$$M = \frac{\widehat{A(t) - A_o}}{A_o} \qquad \text{AM}$$

while for DSB, the modulation index is the ratio of peak instantaneous amplitude to peak unmodulated amplitude

$$M = \hat{A}/A_o \qquad \text{DSB}$$

The result is that with AM, the modulated carrier contains the unmodulated carrier as an additive term, whereas with DSB it does not and the carrier is said to be suppressed. For a given baseband and the same value of modulation index, there is no other difference between the AM and the DSB modulated carriers. However, while the

unmodulated carrier component carries no information, it does endow the AM modulated carrier with a unique property. So long as the modulation index is no greater than 1 (100 percent modulation), the instantaneous amplitude $A(t) = \left[1 + Me_B(t)\right] A_o$ will nearly always be positive* and hence equal to the envelope $|A(t)|$ of the modulated carrier. This insures that a simple envelope demodulator** will yield a virtually undistorted replica of the baseband. However, if a product demodulator** is used, baseband recovery does not depend on the existence of a transmitted carrier component and M may have any value.

PM and FM are also closely related. In each there is a baseband-dependent deviation of phase angle and no amplitude modulation; both are thus pure angle modulation techniques. But where in PM it is the deviation of instantaneous phase from its unmodulated value that is proportional to the baseband, in FM it is the deviation of instantaneous frequency that is proportional. The degree of modulation (modulation index) for PM is specified by the constant of proportionality Φ which has already been identified as the peak phase deviation. The modulation index for FM is given by the ratio of the peak angular frequency deviation Ω to the highest angular frequency in the baseband.

It should be noted that, unless the baseband signal is available for comparison, it is impossible to tell whether an angle modulated

---

*With M=1, A(t) is negative only when $e_B(t) \equiv E_B(t)/\hat{E}_B < -1$ which is permitted to occur only with a small probability.

**To be discussed in detail later in this section.

carrier was produced by PM or FM. The question is decided only by whether it is $\varphi(t)$ or $\dot{\varphi}(t)$ that is proportional to $e_B(t)$. If the former, it is PM; if the latter, FM. As a corollary, Table 1 shows that frequency modulation by $\dot{e}_B(t)$, the time derivative of the baseband, leads to exactly the same modulated waveform as direct phase modulation by $e_B(t)$. Similarly, if the baseband is integrated prior to phase modulation, the result is the same as direct frequency modulation.

Finally, the single-sideband techniques, SSB and SSBFM, are similar to one another in that each requires modulation of both amplitude and angle. Although it is true that the SSB waveform is usually obtained in practice by filtering or otherwise retaining only one sideband from an AM or DSB modulated carrier, its production by direct modulation alone requires that phase (or frequency) be varied simultaneously with amplitude. Indeed, it is the angle modulation that performs the sideband suppression in SSB. In an analogous way, with SSBFM it is the amplitude modulation that suppresses one sideband. Another similarity, noted earlier, is that both SSB and SSBFM involve modulation by the Hilbert transform of the baseband.

The single-sideband methods also exhibit a number of differences. For example, SSBFM cannot be obtained by merely filtering one sideband from an FM waveform. As will be seen shortly, the spectrum of the one sideband in SSBFM is quite different from either of the sidebands in FM. On the other hand, SSBFM can be obtained from an FM modulated carrier by merely superimposing the proper modulation of amplitude whereas to obtain SSB from its double-sideband counterpart,

DSB, requires not only proper modulation of phase but also a change in the character of the amplitude modulation. Notice in particular that the instantaneous amplitude of a SSB modulated carrier is not proportional to the baseband, but rather to a combination of the baseband and its Hilbert transform.

The relation between SSB and DSB is probably better illustrated by rewriting the expression for the SSB modulated carrier in the equivalent but less formidable looking form

$$e(t) = Me_B(t) \cos \omega_o t \mp M\check{e}_B(t) \sin \omega_o t \qquad \text{SSB} \qquad (8)$$

Here, as in Table 1, the upper sign corresponds to an upper sideband and the lower to a lower sideband. From this relation it is apparent that one way of obtaining SSB from DSB is to subtract (or add) to the DSB modulated carrier a quadrature (90^o out of phase) carrier that has been DSB modulated by the Hilbert transform of the baseband. This is the basis of the "phase shift" or "sideband cancellation" technique of SSB modulation.

Modulation as an Operation on the Baseband--Fourier Spectrum Relations

In Table 1, a number of analog modulation methods were defined in terms of simultaneous operations on the amplitude and phase (or frequency) of a sinusoidal carrier. As noted in the introduction to this section, however, an alternate approach is to define modulation as the result of two steps: (1) direct operation on the baseband alone to produce a new signal with particular characteristics in the time and frequency domain, and (2) translation of the spectrum of the new signal to the vicinity of the carrier frequency. The mathematical

where $z^*(t) \triangleq x(t) - jy(t)$ is the complex conjugate of $z(t)$.

In choosing a complex representation for a real waveform $x(t)$, the imaginary part $y(t)$ may be set equal to almost any mathematically convenient real time function, but the most generally useful representation is obtained when

$$y(t) = \breve{x}(t)$$

where $\breve{x}(t)$ is the Hilbert transform of $x(t)$ introduced in Eq. (7). This particular complex representation is called the "analytic signal" representation[4,6,8]* of $x(t)$ and will be denoted $\underline{x}(t)$:

$$\underline{x}(t) = x(t) + j\breve{x}(t) \tag{13}$$

The most familiar examples of analytic signals are the rotating vectors or phasors used almost universally in ac circuit theory to represent simple sinusoidal voltages and currents. Thus, if

$$x_o(t) = C \cos(\omega t + \varphi) \tag{14}$$

it follows from Eq. (7) that

$$\breve{x}_o(t) = C \sin(\omega t + \varphi) \tag{15}$$

The analytic signal representation of $x_o(t)$ is then

$$\underline{x}_o = Ce^{j(\omega t+\varphi)} \tag{16}$$

---

*Also called the "preenvelope"[9] and (when divided by two) the "positive frequency content."[10] The name "analytic signal" reflects the fact that the complex function obtained by evaluating an analytic function of a complex variable $t + jy$ on the real axis $y = 0$ is an analytic signal in the sense of Eq. (13): its real and imaginary parts form a Hilbert transform pair.

description of this approach requires an understanding of complex waveforms and the properties of their Fourier spectra. The necessary background material is presented in some detail in Appendix A and is summarized here for reference.

A complex waveform is a complex-valued function of time

$$z(t) = x(t) + jy(t) = r(t)e^{j\theta(t)} \tag{9}$$

where $x(t)$, $y(t)$, $r(t)$ and $\theta(t)$ are all real-valued time functions related by the conditions

$$r(t) = |z(t)| = \sqrt{x^2(t) + y^2(t)} \tag{10}$$

$$\theta(t) = \arg z(t) = \tan^{-1}[y(t)/x(t)] \tag{11}$$

The waveform $z(t)$ may be interpreted as a time-varying vector* with cartesian components $x(t)$ and $y(t)$, length $r(t)$, and polar angle $\theta(t)$ (measured counterclockwise from the x-axis). The mathematical analysis of a real signal $x(t)$ is often simplified by replacing it with a complex waveform $z(t)$ satisfying the condition

$$x(t) = \text{Re } z(t)$$

The waveform $x(t)$ may then be written in three equivalent ways

$$x(t) = \begin{cases} \text{Re } z(t) \\ \text{Re } z^*(t) \\ \frac{1}{2}[z(t) + z^*(t)] \end{cases} \tag{12}$$

*See Ref. 7 for a provocative defense and extension of the vector interpretation of complex time functions in oscillation analysis.

Geometrically, this is a rotating vector with constant length C, initial angle φ, and uniform counterclockwise angular velocity ω. Corresponding to Eq. (12), $x_o(t)$ can also be written in three equivalent ways:

$$C \cos(\omega t + \varphi) = x_o(t) = \begin{cases} \text{Re } \underline{x}_o(t) \\ \text{Re } \underline{x}_o^*(t) \\ \frac{1}{2}\left[\underline{x}_o(t) + \underline{x}_o^*(t)\right] \end{cases} \tag{17}$$

Rotating vectors also find extensive use in the description of Fourier spectra. The concept of a Fourier spectrum is based on the fact that a nonperiodic real waveform x(t) with finite total energy can be written as an integral over real sinusoids whose angular frequencies $\omega = 2\pi f$ lie in the continuum $0 \le \omega \le \infty$:

$$x(t) = \frac{1}{\pi}\int_0^\infty C(\omega) \cos\left[\omega t + \varphi(\omega)\right] d\omega \tag{18}$$

The amplitude $C(\omega)$ and phase $\varphi(\omega)$ of the component sinusoids may be computed from x(t) by means of the formulas:

$$\left.\begin{aligned} C(\omega) &= \sqrt{A^2(\omega) + B^2(\omega)}\,, \qquad \varphi(\omega) = \tan^{-1}\left[B(\omega)/A(\omega)\right] \\ &\text{where} \\ A(\omega) &= \int_{-\infty}^{\infty} x(t) \cos \omega t \, dt\,, \qquad B(\omega) = \int_{-\infty}^{\infty} x(t) \sin \omega t \, dt \end{aligned}\right\} \tag{19}$$

If each sinusoid in the integrand of Eq. (18) is now replaced by its analytic signal representation, i.e., the counterclockwise rotating vector $X(\omega)e^{j\omega t}$ where

$$X(\omega) \equiv C(\omega)e^{j\varphi(\omega)} \tag{20}$$

is the value of the rotating vector at time zero (its initial value), the result is the analytic signal representation of $x(t)$:

$$\underline{x}(t) = \frac{1}{\pi}\int_0^\infty X(\omega)e^{j\omega t}\,d\omega \tag{21}$$

The analytic signal representation of a real waveform may thus be interpreted as the rotating vector of variable length and angular velocity obtained by summing the rotating vectors which represent its Fourier component sinusoids.

Similarly, if each sinusoid is replaced by the complex conjugate of its analytic signal representation, namely, the clockwise rotating vector $X^*(\omega)e^{-j\omega t}$, the result is the conjugate of the analytic signal representation of $x(t)$:

$$\underline{x}^*(t) = \frac{1}{\pi}\int_{-\infty}^0 X(\omega)e^{j\omega t}\,d\omega \tag{22}$$

Finally, if each sinusoid is replaced by the sum of counter-rotating vectors, as in the third line of Eq. (17), the result is the so-called Fourier integral representation of $x(t)$:

$$x(t) = \frac{1}{2}\left[\underline{x}(t) + \underline{x}^*(t)\right] = \frac{1}{2\pi}\int_{-\infty}^\infty X(\omega)e^{j\omega t}\,d\omega \tag{23}$$

The Fourier integral thus expresses the real waveform $x(t)$ as the sum (integral) of a "spectrum" of complex waveforms, viz., an infinite set of rotating vectors having both positive (counterclockwise) and negative (clockwise) angular velocities. The function $X(\omega)$ which specifies these rotating vectors in terms of their initial values is called the "Fourier spectrum." The spectrum $X(\omega)$ is also called the Fourier transform* of $x(t)$ because, with the aid of Eqs. (19) and (20), it can be expressed as an integral operation $\mathfrak{F}$ on $x(t)$:

$$X(\omega) = \mathfrak{F}[x(t)] = \int_{-\infty}^{\infty} x(t)e^{-j\omega t}\, dt \qquad (24)$$

From Eq. (20), it is clear that the modulus (length) and argument (polar angle) of the spectrum $X(\omega)$ respectively give the amplitude $C(\omega)$ and initial phase $\varphi(\omega)$ of the real sinusoid with frequency $f = |\omega/2\pi|$ in the trigonometric Fourier integral of Eq. (18). Because of the close connection between vector angular velocity ω and sinusoidal frequency f, counterclockwise and clockwise values of ω are commonly referred to respectively as positive and negative (angular) frequencies when describing spectra.

The Fourier integral representation of Eq. (23) generalizes readily to complex waveforms without change in form:

$$z(t) = x(t) + jy(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Z(\omega)e^{j\omega t}\, d\omega \qquad (25)$$

where the spectrum or Fourier transform of $z(t)$ is given by

$$Z(\omega) = X(\omega) + jY(\omega) = \int_{-\infty}^{\infty} z(t)e^{-j\omega t}\, dt \qquad (26)$$

*The Fourier transform will usually be denoted by the upper case version of the letter used for the corresponding time function.

It is obvious from Eq. (25) that the spectrum of the complex conjugate of $z(t)$ is the complex conjugate of the spectrum of $z(t)$, but with positive and negative frequencies interchanged:

$$\mathfrak{F}[z^*(t)] = Z^*(-\omega) \tag{27}$$

With the aid of the foregoing relations, the Fourier spectra of the various types of waveforms that are of interest in modulation theory may be described as follows:

- The spectrum of a real waveform such as $x(t)$ is two-sided (contains both positive and negative frequencies) and exhibits conjugate or Hermitian symmetry:

$$X(-\omega) = X^*(\omega) \tag{28}$$

- The spectrum of an arbitrary complex waveform such as that in Eq. (9) is two-sided and in general unsymmetrical. Its properties depend on the spectra $X(\omega)$ and $Y(\omega)$ of its real and imaginary parts.
- The spectrum $\underline{X}(\omega)$ of the analytic signal $\underline{x}(t)$ is one-sided and contains no negative frequencies. It is equal to the positive frequency spectrum of $2x(t)$:

$$\mathfrak{F}\ [\underline{x}(t)] = \underline{X}(\omega) = [1 + \operatorname{sgn}(\omega)]\ X(\omega) \tag{29}$$

where

$$\operatorname{sgn}(\omega) = \begin{cases} -1, & \omega < 0 \\ 0, & \omega = 0 \\ 1, & \omega > 0 \end{cases} \tag{30}$$

- The spectrum of $\underline{x}^*(t)$, the complex conjugate of the analytic signal representation of $x(t)$, is also one-sided, but here it is the positive frequencies that vanish:

$$\mathfrak{F}\left[\underline{x}^*(t)\right] = \underline{X}^*(-\omega) = \left[1 - \mathrm{sgn}(\omega)\right] X(\omega) \tag{31}$$

- The spectrum of the Hilbert transform of $x(t)$ may be found by combining Eqs. (13) and (29):

$$\mathfrak{F}\left[\breve{x}(t)\right] = -j\left[\underline{X}(\omega) - X(\omega)\right] = -j\ \mathrm{sgn}(\omega)\ X(\omega) \tag{32}$$

  This leads to the interpretation of $\breve{x}(t)$ as the harmonic conjugate of $x(t)$, i.e., a waveform whose spectral components lag those of $x(t)$ by 90^o at all frequencies.

- The spectrum of the product of a complex waveform $m(t)$ and the rotating vector $e^{j\omega_o t}$ is just the spectrum of $m(t)$ shifted upward in angular frequency by ω_o

$$\mathfrak{F}\left[m(t)e^{j\omega_o t}\right] = M(\omega - \omega_o) \tag{33}$$

- A complex waveform whose negative frequency spectrum vanishes is the analytic signal representation of its real part. Likewise, a complex waveform whose spectrum contains no positive frequencies is the complex conjugate of the analytic signal representation of its real part.

Returning now to the description of modulation as an operation on the baseband rather than on the carrier,* it is only necessary to

*The approach used here is similar to that of Bedrosian.(4) Voelcker(8) has recently described an even more general approach to modulation theory.

note that a complex representation c(t) of any real modulated waveform

$$e(t) = a(t) \cos [\omega_o t + \varphi(t)]$$

can be obtained by forming the product

$$c(t) = m(t)\ \underline{e}_o(t) \tag{34}$$

where

$$m(t) \equiv a(t)\ e^{j\varphi(t)}$$

is a complex waveform derived from the baseband, and

$$\underline{e}_o(t) \equiv e^{j\omega_o t}$$

is the analytic signal (rotating vector) representation of the normalized sinusoidal carrier $\cos \omega_o t$.

Thus, if modulation is viewed as the process of generating the complex modulated carrier c(t), the first step in the process is the operation on the baseband which yields m(t), and the second is the multiplication of m(t) by $\underline{e}_o(t)$. By Eq. (33), the latter step merely serves to translate the spectrum M(ω) of m(t) up to the carrier frequency ω_o:

$$C(\omega) = M(\omega - \omega_o) \tag{35}$$

Since this second step is the same for all modulation methods, a given method is characterized entirely by the function m(t) which is derived from the baseband $e_B(t)$ in the first step. For this reason, m(t) is called the "modulation function."

It is of course obvious that the only modulation functions of practical interest are those which yield modulated carriers

$$e(t) = \text{Re } c(t) = a(t) \cos [\omega_o t + \varphi(t)] \tag{36}$$

from which the baseband, or some simple function of it, can be recovered by one of the basic types of demodulator discussed later in this section: i.e., product, envelope, phase and frequency. In that discussion, the responses of these demodulators to the signal e(t) are shown to be proportional respectively to the following parameters of e(t):

$$\left.\begin{aligned} &\text{In-phase signal component} = a'(t) \cos \varphi'(t) \\ &\text{Envelope} = a'(t) \\ &\text{Phase deviation} = \varphi'(t) \\ &\text{Frequency deviation} = \dot{\varphi}'(t) \end{aligned}\right\} \tag{37}$$

Here, as in Appendix A, $a'(t)$ and $\varphi'(t)$ are defined in terms of the modulus and argument of the analytic signal representation $\underline{e}(t)$ of e(t):

$$a'(t) = |\underline{e}(t)| = \sqrt{e^2(t) + \breve{e}^2(t)}$$

$$\omega_o t + \varphi'(t) = \arg \underline{e}(t) = \tan^{-1}[\breve{e}(t)/e(t)]$$

where

$$\underline{e}(t) = e(t) + j\breve{e}(t) = a'(t)\, e^{j\varphi'(t)}\, e^{j\omega_o t}$$

Before going further, it is important to note that, although

$$e(t) = \text{Re } \underline{e}(t) = a'(t) \cos [\omega_o t + \varphi'(t)]$$

has exactly the same form as Eq. (36), it cannot be concluded that $c(t)$ is equal to $\underline{e}(t)$. And unless this is true, the envelope $a'(t)$ and phase deviation $\varphi'(t)$ which characterize the demodulator outputs will not be equal respectively to the waveforms $a(t)$ and $\varphi(t)$ which specify the modulation. Under these circumstances, it is not in general possible to determine whether a given modulation function will yield a useful modulated carrier. Even if the applied modulations $a(t)$ and/or $\varphi(t)$ have a simple dependence on the baseband, there is no assurance that this will be true of the demodulator outputs.

To avoid this difficulty and to make it possible to test an arbitrary modulation function for detectability, it is necessary first to impose the condition that its negative frequency spectrum not extend below $-\omega_o$, i.e.,

$$M(\omega) = 0 \quad \text{for} \quad \omega < -\omega_o \tag{38}$$

If $m(t)$ meets this condition, Eq. (35) shows that the complex carrier $c(t)$ will contain no negative frequencies and so will be the analytic signal representation of its real part $e(t)$. Thus, Eq. (38) implies

$$c(t) = \underline{e}(t) \tag{39}$$

and hence

$$a'(t) = a(t) \equiv |m(t)|, \qquad \varphi'(t) = \varphi(t) \equiv \arg m(t)$$

Combining this with Eq. (37), the conditions on $m(t)$ for recovery of a simple function $b(t)$ of the baseband by product, envelope, phase, and frequency demodulators are then respectively:

$$\text{Re } m(t) \propto b(t) \qquad \text{Product} \tag{40a}$$

$$|m(t)| \propto b(t) \qquad \text{Envelope} \tag{40b}$$

$$\arg m(t) \propto b(t) \qquad \text{Phase} \tag{40c}$$

$$\arg m(t) \propto \int b(t)\, dt \qquad \text{Frequency} \tag{40d}$$

Any modulation function m(t) which makes c(t) an analytic signal and satisfies one of the conditions just stated will define a modulated carrier from which the baseband function b(t) can be recovered.

The conditions on m(t) given by Eqs. (38) and (40) are not excessively restrictive. They allow considerable latitude for specifying the functional dependence of m(t) on b(t), which is another way of saying that they permit the description of many different modulation methods. In particular, for any given type of demodulator, it is possible to define modulated carriers having a wide variety of spectral characteristics. The spectrum E(ω) of the normalized modulated carrier may in fact be expressed directly in terms of the spectrum M(ω) of the modulation function. Taking the Fourier transform of the waveform relation

$$e(t) = \text{Re } c(t) = \frac{1}{2}\,[c(t) + c^*(t)]$$

it follows from Eq. (27) that

$$E(\omega) = \frac{1}{2}\,[C(\omega) + C^*(-\omega)]$$

and so by Eq. (35),

$$E(\omega) = \frac{1}{2}\,[M(\omega - \omega_o) + M^*(-\omega - \omega_o)] \tag{41}$$

Due to the condition expressed by Eq. (38), the spectra represented by the two terms on the right-hand side of Eq. (41) never overlap. The first term represents the positive frequency part of $E(\omega)$, the second its negative frequency part. In the special case where $m(t)$ is real, $M^*(-\omega) = M(\omega)$, and

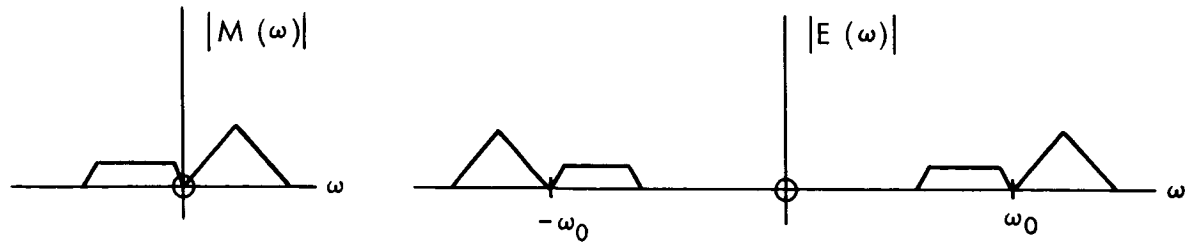
$$E(\omega) = \frac{1}{2}\left[M(\omega - \omega_o) + M(\omega + \omega_o)\right] \qquad \text{real } m(t) \qquad (42)$$

In all cases, the spectrum of $e(t)$ may be found by shifting the entire spectrum of $m(t)$ upward in frequency through ω_o to obtain the positive frequency part, then reflecting it about zero frequency and taking its conjugate to find the negative frequency part.

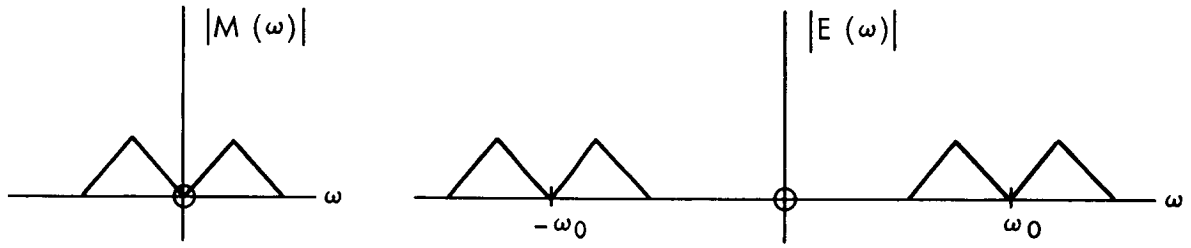
From Eq. (41) or (42), it is clear that when $m(t)$ contains both positive and negative frequencies, $e(t)$ will be double sidebanded, i.e., its spectral components will lie on both sides of the carrier frequency (Fig. 2a); if $m(t)$ is also real, the sidebands will be symmetric about ω_o (Fig. 2b). On the other hand, if $m(t)$ is an analytic signal, it will have only positive frequencies and the carrier will have only an upper sideband (Fig. 2c). Similarly if $m(t)$ is the complex conjugate of an analytic signal, $e(t)$ will have only a lower sideband (Fig. 2d). As an aid in constructing such one-sided modulation functions, it is useful to note that an analytic function of an analytic signal is itself an analytic signal.

To illustrate the application of these relationships, consider the family of modulation functions which leads to carriers which may be demodulated by a product demodulator. This family is defined by the conditions of Eqs. (38) and (40a) and hence $m(t)$ has the form

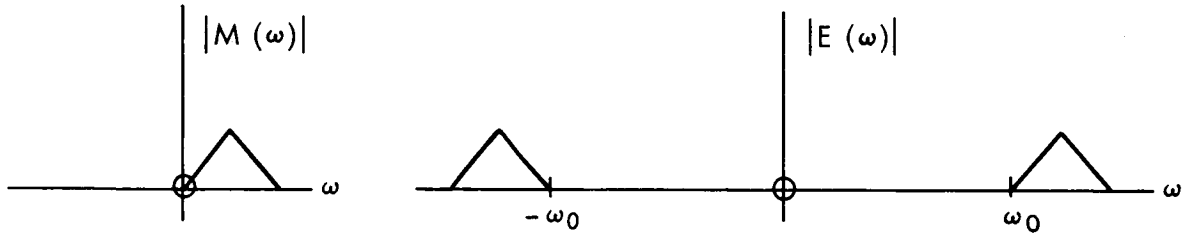
$$m(t) = b(t) + jy(t) \qquad (43)$$



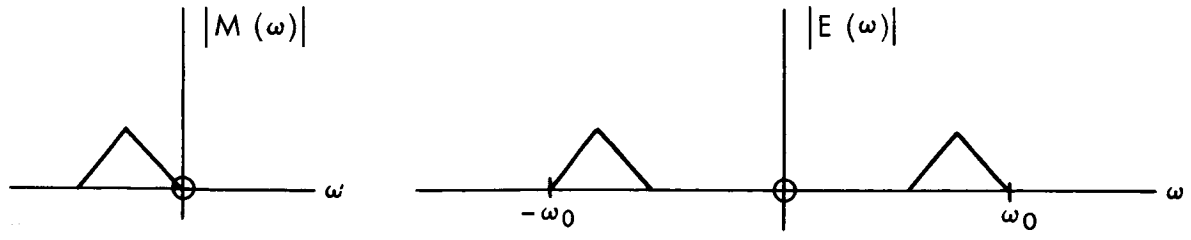
a. Modulation function is complex but arbitrary



b. Modulation function is real



c. Modulation function is an analytic signal



d. Modulation function is the complex conjugate of an analytic signal

Fig. 2—Relation between spectra of modulation function and modulated carrier

where b(t) is the baseband function to be recovered and y(t) is a real waveform to be chosen arbitrarily, each subject only to the condition that its spectrum is confined to the range $\omega > -\omega_o$. Setting

$$b(t) = M\, e_B(t)$$

there are two cases of special interest:

1. $\underline{y(t) = 0}$. In this case m(t) is a real waveform

$$m(t) = M\, e_B(t)$$

with spectrum

$$M(\omega) = M\, E_B(\omega)$$

where $E_B(\omega)$ is the spectrum of the normalized baseband signal.* The normalized modulated carrier defined by m(t) is

$$e(t) \equiv \text{Re}\,[m(t)\, e^{j\omega_o t}] = M\, e_B(t)\, \cos\, \omega_o t$$

and using Eq. (42), its spectrum is

$$E(\omega) = \tfrac{1}{2}\, M[E_B(\omega - \omega_o) + E_B(\omega + \omega_o)] \qquad (44)$$

With a symmetrical double-sidebanded spectrum and no carrier component, this case obviously corresponds to DSB modulation.

2. $\underline{y(t) = \breve{b}(t)}$. In this case, m(t) is the analytic signal representation of b(t)

$$m(t) = b(t) + j\breve{b}(t) \equiv \underline{b}(t) = M\, \underline{e}_B(t)$$

and by Eq. (29) has a one-sided spectrum with only positive frequencies

$$M(\omega) = M\underline{E}_B(\omega) = M[1 + \text{sgn}\,(\omega)]\, E_B(\omega)$$

---

*$E_B(\omega)$ should not be confused with the unnormalized baseband waveform $E_B(t)$.

The modulated carrier is

$$e(t) = M\, e_B(t) \cos \omega_o t - M\, \breve{e}_B(t) \sin \omega_o t$$

and by Eq. (41) its spectrum is

$$E(\omega) = \frac{M}{2} \left\{ [1 + \operatorname{sgn} (\omega - \omega_o)]\, E_B(\omega - \omega_o) + [1 - \operatorname{sgn} (\omega + \omega_o)]\, E_B (\omega + \omega_o) \right\} \tag{45}$$

This expression has a simple interpretation. It differs from the DSB spectrum of case 1 given by Eq. (44) only in the removal of the lower sideband. The modulation involved is thus SSB. The corresponding lower-sideband version of SSB can be obtained by taking

$$y(t) = - \breve{b}(t)$$

to obtain the complex conjugate modulation function

$$m(t) = \underline{b}^*(t) = M\, \underline{e}_B^*(t)$$

A second major family of modulation functions is that corresponding to the conditions Eqs. (38) and (40c) for detectability by a phase demodulator. To meet Eq. (40c), the members of this family must have the form

$$m(t) = y(t)\, e^{jb(t)} \tag{46}$$

where again $b(t)$ is a function of the baseband and $y(t)$ is an arbitrary real waveform. To satisfy Eq. (38), $b(t)$ and $y(t)$ must be such that the convolution of the spectra of $y(t)$ and $\exp[jb(t)]$ vanishes for $\omega < -\omega_o$. Setting

$$b(t) = \Phi\, e_B(t)$$

there are two cases of special interest.

1. $\underline{y(t) = 1}$. In this case

$$m(t) = e^{jb(t)} = e^{j\Phi e_B(t)}$$

and

$$e(t) = \cos\,[\omega_o t + \Phi e_B(t)]$$

The spectrum $E(\omega)$ of $e(t)$ can be expressed in terms of $M(\omega)$ by Eq. (41), but due to the nonlinear relation between $m(t)$ and $e_B(t)$, it cannot be simply expressed in terms of the baseband spectrum $E_B(\omega)$. Since $m(t)$ is neither real nor an analytic signal, the sidebands of $e(t)$ will be double and unsymmetrical. From the form of $e(t)$ the modulation is PM.

2. $\underline{y(t) = e^{-\breve{b}(t)}}$. In this case

$$m(t) = e^{j[b(t) + j\breve{b}(t)]} = e^{j\underline{b}(t)} = e^{j\Phi\underline{e}_B(t)}$$

and, as an analytic function of an analytic signal, is itself an analytic signal. Hence the spectrum of the phase modulated carrier

$$e(t) = e^{-\Phi\breve{e}_B(t)}\cos\,[\omega_o t + \Phi e_B(t)]$$

will have only an upper sideband. Accordingly, the modulation method is called single-sideband phase modulation (SSBPM). To obtain the lower-sideband version of SSBPM, the complex conjugate modulation function is used.

The foregoing examples for the product and phase demodulator make it apparent that the modulation function approach provides a systematic method for defining modulated carriers with prescribed demodulation and spectral properties. Indeed, with proper choices for the baseband function b(t), it is possible to describe nearly every modulation method of practical interest using modulation functions whose dependence on the baseband is given by either a linear or an exponential function of b(t). This fact permits a classification of modulation techniques that is somewhat more general than the historic distinction between amplitude and angle techniques. Thus a modulation method will be called linear if the dependence of m(t) on b(t) is linear, as for example in Eq. (43), and exponential if the dependence is exponential, as in Eq. (46), for example. The two classes of modulation methods thus distinguished may then be subdivided by the nature of the carrier spectrum (double- or single-sideband) and by the type of demodulator required. Table 2 illustrates this classification method for all of the analog modulation methods discussed in this Memorandum. By using more complicated baseband functions, it may be extended to include modulation methods whose carriers can be demodulated by both AM and SSB receivers.[4] Such methods are not discussed here because in satellite applications they offer no special advantages over more conventional techniques.

Despite their great utility in theoretical analyses, Fourier spectra are not well suited to the description of the modulated carriers most commonly encountered in practice because a Fourier spectrum can represent only a single explicit or deterministic waveform.

Table 2

DEFINITION OF ANALOG MODULATION METHODS IN TERMS OF OPERATIONS ON THE BASEBAND

Classification		Modulation Function $m(t)$	Baseband Function $b(t)$	Normalized Modulated Carrier $e(t) = \mathrm{Re}[m(t)e^{j\omega_o t}]$	Usual Designation
Linear	Double sideband	$b(t)$	$1 + Me_B(t)$	$[1 + Me_B(t)]\cos \omega_o t$	AM
			$Me_B(t)$	$Me_B(t)\cos \omega_o t$	DSB
	Single sideband[a]	$\underline{b}(t)$ or $\underline{b}^*(t)$	$Me_B(t)$	$Me_B(t)\cos \omega_o t \mp M\check{e}_B(t)\sin \omega_o t$	SSB
Exponential	Double sideband	$e^{jb(t)}$	$\Phi e_B(t)$	$\cos[\omega_o t + \Phi e_B(t)]$	PM
			$\Omega \int e_B(t)dt$	$\cos[\omega_o t + \Omega \int e_B(t)dt]$	FM
	Single sideband[a]	$e^{j\underline{b}(t)}$ or $e^{j\underline{b}^*(t)}$	$\Phi e_B(t)$	$e^{\mp \Phi \check{e}_B(t)} \cos[\omega_o t + \Phi e_B(t)]$	SSBPM
			$\Omega \int e_B(t)dt$	$e^{\mp \Omega \int \check{e}_B(t)dt} \cdot \cos[\omega_o t + \Omega \int e_B(t)dt]$	SSBFM

[a]Where alternate signs appear, the upper sign corresponds to the use of the first modulation function and to a carrier with an upper sideband. The lower sign corresponds to the use of the second modulation function and a carrier with a lower sideband.

An information-bearing carrier, on the other hand, has an inherent unpredictability, or randomness, that cannot be represented by a deterministic waveform. About the only basebands of practical importance which have well-defined Fourier spectra are the periodic test signals used for signalling, trouble shooting, and system adjustment. Examples of these spectra will be calculated later in connection with the discussion of the "power spectrum," which is the spectral concept appropriate to the description of random waveforms.

Statistical Concepts

To this point, an analog modulated sinusoidal carrier has been described only to the extent of showing how its normalized waveform $e(t)$ and Fourier spectrum $E(\omega)$ depend on the corresponding baseband waveform $e_B(t)$ and spectrum $E_B(\omega)$. It was noted that this description is exact providing that the waveforms of the baseband and the unmodulated carrier are known exactly. It was also noted that, except for test signals, the basebands of practical interest cannot be described by explicit or deterministic waveforms. Indeed, in order to carry real messages or "information," the baseband must be intrinsically unpredictable and can only be described in statistical terms. The problem then is to derive a statistical description of the modulated carrier, given a statistical description of the baseband and the functional relation between the two waveforms as summarized, for example, in Tables 1 and 2.

Although the statistical properties of an information-bearing signal cannot in general be represented by a single waveform of finite

duration, they can be described in terms of a mathematical model consisting of an infinite set, or "ensemble," of explicit waveforms, each waveform corresponding to one of the possible messages that the signal could carry. The individual waveforms are called "sample functions," and the ensemble of sample functions is called a "random process."[6,10] The statistics in question are then determined at a specified time by the properties of the random process at that time.

In applications to communication systems, however, it is normally assumed that, although each sample function in the ensemble is time dependent, the properties of the ensemble as a whole are not, i.e., the ensemble is "stationary." In addition, it will be assumed here that the random process is "ergodic," which means that all of the sample functions have the same statistics. As a consequence, "fraction-of-ensemble" statistics at a particular time are the same as the "fraction-of-time" statistics of a particular sample function. Put another way, averages over the ensemble at a particular time are equal to averages over time for a particular sample function. The motivation for the ergodic assumption is that it "legalizes" the practical necessity of inferring the properties of the statistical model--the ensemble--from experimental fraction-of-time measurements on samples of the actual signal that the model is to represent.

For most engineering purposes, the "statistics" of a random signal (or more accurately, the random process that represents it) are adequately described by two mathematical functions called the "probability density function" and the "power spectral density," or

simply "power spectrum."* Once these density functions are known for a random signal, they can be used to answer many relevant engineering questions about it. In particular, they can be used respectively to compute the ratio of peak-to-average signal power, and the amount of spectrum or bandwidth which the signal occupies. These two quantities provide a summary description of the signal that is quite useful in designing the transmitting and receiving amplifiers that must handle it.

The probability density function for a random signal E(t) is a real nonnegative function $p_E(x)$ such that $p_E(x)dx$ is the probability that the instantaneous value of the signal will lie in the infinitesimal range x to x + dx. By virtue of the ergodic hypothesis, this probability can be interpreted as the fraction of time that E(t) has values in this range.

Thus, the probability density function can be used to compute the time average of a function f[E(t)] of signal level. It is only necessary to weight f(x) with its probability of occurrence $p_E(x)dx$ and then integrate over all signal levels:**

---

*A random process may equally well be described by the Fourier transforms of its probability density function and power spectrum. These are called respectively the "characteristic function" and the "autocorrelation function" and will be introduced later in this section.

**Different names ("mean," "expectation") and notations (E[f(E)], $\langle f(E) \rangle$) are sometimes applied to the statistical or ensemble average on the right-hand side of this equation to distinguish it from the time average on the left-hand side. The ergodic assumption makes such distinctions unnecessary in this report.

$$\overline{f[E(t)]} \equiv \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f[E(t)]\, dt = \int_{-\infty}^{\infty} f(x)\ p_E(x)\, dx \tag{47}$$

For example, the average value of the signal itself is

$$\overline{E(t)} = \int_{-\infty}^{\infty} x\ p_E(x)\, dx$$

and the average value of the square of the signal is

$$S \equiv \overline{E^2(t)} = \int_{-\infty}^{\infty} x^2\ p_E(x)\, dx \tag{48}$$

When E(t) is a current or voltage, the quantity S here defined is the average signal power absorbed in a one-ohm load.* Since the average value $\overline{E(t)}$ of most signals of interest in communication engineering is zero, the mean power S is also equal to the variance σ_E^2 of the signal, defined as the mean square deviation of the signal from its average value:

$$\sigma_E^2 \equiv \overline{[E(t) - \overline{E(t)}]^2} \equiv \overline{E^2(t)} - \overline{E(t)}^2 \tag{49}$$

From its definition, it is clear that the probability density function can also be used to compute the probability, or fraction of the time, that a function of signal level f[E(t)] occupies a specified range of values. Conversely, it can be used to determine the value which the function exceeds with a specified probability. In particular, $p_E(x)$ can be used to calculate the peak signal power, defined as the value $\hat{S}$ which the instantaneous power $E^2(t)$ exceeds with only a specified

*In speaking of power, a reference load of one ohm will consistently be assumed.

small probability ϵ. If ϵ is the probability (or fraction of the time) that $E^2(t)$ exceeds $\hat{S}$, then $1-\epsilon$ is the probability that it does not. Since the latter is identical to the probability that E(t) lies in the range $\left(-\sqrt{\hat{S}}, \sqrt{\hat{S}}\right)$, the peak power is given in terms of ϵ and $p_E(x)$ by

$$\int_{-\sqrt{\hat{S}}}^{\sqrt{\hat{S}}} p_E(x)dx = 1 - \epsilon \tag{50}$$

The peak power $\hat{S}$ is thus a function of ϵ, and when ϵ is very small $\hat{S}$ measures the extreme deviations or peaks of the signal in somewhat the same way that the average power S measures the average magnitude of the positive and negative excursions of the signal from its mean value. The ratio of peak to average power

$$\Lambda \equiv \hat{S}/S$$

forms a convenient dimensionless index of signal "peakiness" that is independent of the absolute signal level.

The importance of the concepts of average and peak power to system design is well illustrated by the case of the modulated carrier signal. To begin with, as will be shown in the discussion of demodulators, the average power S_R at the receiver input determines the SNR at the receiver output, and hence the quality of the output signal. Moreover, when the value of S_R which yields the desired output signal quality is known, the average transmitter output power $S_T = \overline{E_T^2(t)}$ can be calculated from the relation

$$S_T = \frac{L}{G_T G_R} S_R \tag{51}$$

where L is the path loss and G_T and G_R are the gains of the transmitting and receiving antennas, respectively.

The peak received carrier power $\hat{S}_R$ is important because the value of the probability ϵ involved in its definition is chosen on the basis of experimental measurements of the distortion produced by nonlinearities in the amplifiers through which the modulated carrier must pass. In particular, ϵ is chosen so that, with $E_R(t)$ confined to the range $\pm\sqrt{\hat{S}_R}$ with probability $1-\epsilon$, the distortion will be acceptably low. Thus $\hat{S}_R$ is a measure of the input signal range over which the receiver amplifiers should be linear. In addition, it determines the peak power

$$\hat{S}_T = \frac{L}{G_T G_R} \hat{S}_R \tag{52}$$

which the transmitter power amplifier must handle without excessive distortion. Peak transmitter power can be critically important in satellite repeaters since it often determines the size, weight, and primary power requirement of the transmitter.

The other basic statistical density, the "power spectrum" of a signal E(t), is defined as a function $W_E(f)$ such that $W_E(f)df$ is that portion of the total average power S of the signal contained in the frequency range f to f+df. Thus, where the probability density function could be used to describe the fraction of the time that the instantaneous signal power lies in a specified power interval, the

power spectrum describes the fraction of the average power that lies in a specified frequency interval. In general, the two density functions are independent; i.e., a knowledge of the probability density function conveys no information about the spectral properties of the signal, and vice versa--except for the obvious condition that they must yield the same total signal power:

$$S = \int_{-\infty}^{\infty} W_E(f)df = \int_{-\infty}^{\infty} x^2 p_E(x)dx \tag{53}$$

Since the power spectrum gives the power in an infinitesimal band of frequencies, and the Fourier spectrum specifies the amplitude of the sinusoidal components in such a band, it might be thought that the power spectrum could be expressed in terms of the Fourier spectrum for those deterministic signals that possess a Fourier spectrum. This is in fact true, but only for periodic waveforms. If Z_n is the Fourier spectrum for the (possibly complex) periodic waveform $z(t)$ with period T, then its power spectrum is[11]

$$W_z(f) = \sum_{n=-\infty}^{\infty} |Z_n|^2 \, \delta(f-nf_1), \qquad f_1 = T^{-1} \tag{54}$$

where $\delta(f)$ is the Dirac delta function defined by the conditions that for an arbitrary function $g(f)$ and a constant a,

$$\begin{aligned} &\delta(f - a) = 0, \qquad f \neq a \\ &\int_{-\infty}^{\infty} g(f)\, \delta(f - a)\, df = g(a) \end{aligned} \tag{55}$$

When the waveform is not periodic, it may have a Fourier spectrum and no power spectrum, or it may have a power spectrum and no Fourier spectrum. For example, a nonperiodic deterministic waveform $z(t)$ will have a Fourier spectrum if its total energy $\int_{-\infty}^{\infty} |z(t)|^2 \, dt$ is finite, but in this case, the average power $S = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} |z(t)|^2 \, dt$ vanishes and so the power spectrum is zero.

On the other hand, if the waveform has a non-zero but finite average power, it will have a power spectrum but will in general have no Fourier spectrum. In this case, however, the power spectrum can be expressed in terms of a limiting operation on the Fourier spectrum

$$Z_T(f) = \int_{-\infty}^{\infty} z_T(t) \, e^{-j2\pi ft} \, dt$$

of a finite length sample or truncated version of $z(t)$

$$z_T(t) = \begin{cases} z(t) & |t| < T \\ 0 & |t| > T \end{cases}$$

The power spectrum of $z(t)$ is then given by[10]

$$\int_{f_1}^{f_2} W_z(f) \, df = \lim_{T\to\infty} \frac{1}{2T} \int_{f_1}^{f_2} |Z_T(f)|^2 \, df$$

where f_1 and f_2 are arbitrary frequencies. When applied to a small frequency interval, this definition reflects the well-known fact that the average power of a sinusoid is one-half the square of its amplitude.

A much more useful alternative definition for the power spectrum of an explicit, finite-power waveform z(t) is obtained by first introducing the autocorrelation function of z(t)

$$R_z(\tau) = \overline{z^*(t)z(t+\tau)} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} z^*(t)z(t+\tau)dt \tag{56}$$

Then, by a theorem of Wiener and Khinchin, the power spectrum is just the Fourier spectrum (or transform) of $R_z(\tau)$:

$$W_z(f) = \mathfrak{F}\left[R_z(\tau)\right] = \int_{-\infty}^{\infty} R_z(\tau)\ e^{-j2\pi f\tau}\ d\tau \tag{57}$$

In other words, the power spectrum defines the initial values of the rotating vectors in the Fourier integral representation of $R_z(\tau)$:

$$R_z(\tau) = \int_{-\infty}^{\infty} W_z(f)\ e^{+j2\pi f\tau} df \tag{58}$$

Note from either Eq. (56) or Eq. (58) that the autocorrelation function evaluated for $\tau = 0$ is just the total average power

$$R_z(0) = S \tag{59}$$

Given this procedure for finding the power spectrum of an arbitrary deterministic waveform, the power spectrum of a random waveform (process) may be defined as the ensemble average of the power spectra of the sample functions which describe the process. In this connection, however, it should be noted that if the random process is ergodic, as here assumed, all sample functions have the same autocorrelation functions and power spectra. The ergodic hypothesis thus "justifies" the use of measurements on the random physical signal to infer the power spectrum of the random process which represents it.

Several symmetry properties of autocorrelation functions and power spectra follow directly from the definitions just given. From Eq. (56), the autocorrelation function always exhibits conjugate symmetry

$$R_z(-\tau) = R_z^*(\tau) \tag{60a}$$

and this insures that $W_z(f)$ will be real. If the waveform $z(t)$ is real, then $R_z(\tau)$ is also real, and by Eq. (60a) is symmetric about $\tau = 0$:

$$R_z(-\tau) = R_z(\tau) \quad , \quad z(t) \text{ real} \tag{60b}$$

With both $R_z(\tau)$ and $W_z(f)$ real, it follows that $W_z(f)$ is symmetric about $f = 0$

$$W_z(-f) = W_z(f) \quad , \qquad z(t) \text{ real} \tag{61}$$

The following easily demonstrated[11] relations between power spectra are also of frequent use. The power spectrum of the derivative $\dot{z}(t) = dz(t)/dt$ of a waveform $z(t)$ may be written in terms of the power spectrum $W_z(f)$ of $z(t)$

$$W_{\dot{z}}(f) = \omega^2 W_z(f) \qquad \omega = 2\pi f \tag{62a}$$

Similarly, the power spectrum of the integral $\int z(t)dt$ is given by

$$W_{\int z dt}(f) = \omega^{-2} W_z(f) \tag{62b}$$

If the waveform $z(t)$ with power spectrum $W_z(f)$ is applied to a filter with transfer function $H(f)$, the power spectrum of the output $z_o(t)$ is

$$W_{z_o}(f) = |H(f)|^2 \, W_z(f) \tag{62c}$$

As noted earlier, the power spectrum of a signal can be used to compute the frequency interval or bandwidth which the signal occupies. In speaking of frequency intervals, however, it should be recognized from Eq. (58) that, like the Fourier spectrum, the power spectrum is defined for both positive and negative frequencies.* The frequencies of interest in defining bandwidth, however, are "physical frequencies"--the number of oscillations per second executed by the real sinusoidal components of the signal. In terms of the power spectrum $W_E(f)$ then, the power $S(f_1,f_2)$ corresponding to the physical frequency interval f_1 to f_2 includes the contributions of both positive and negative values of the frequency variable f in the range $f_1 \leq |f| \leq f_2$:

$$S(f_1,f_2) = \int_{-f_1}^{-f_2} W_E(f)df + \int_{f_1}^{f_2} W_E(f)df$$

In the case of real signals, such as the baseband $E_B(t)$ and received modulated carrier $E_R(t)$, Eq. (61) shows that this may be written as a single integral over positive frequencies

$$S(f_1,f_2) = 2\int_{f_1}^{f_2} W_E(f)df$$

Not surprisingly, there is no single universally appropriate definition for the bandwidth of a signal. For some applications it is convenient

*Again, the term "negative frequencies" is just a shorthand way of referring to the angular velocities of the clockwise rotating vectors in a Fourier integral, in this case the Fourier integral for the autocorrelation function.

to observe that the normalized power spectrum

$$w_E(f) \equiv W_E(f)/S \equiv \frac{W_E(f)}{\int_{-\infty}^{\infty} W_E(f)\,df} \tag{63}$$

has many of the properties of a probability density function in which signal level has been replaced by frequency. For real, low-pass* signals, $w_E(f)$ is symmetrical about and concentrated near zero frequency, and a measure of the mean extent of the power spectrum is the rms bandwidth, defined by analogy with the standard deviation of a random variable by

$$B^2 = \int_{-\infty}^{\infty} f^2\, w_E(f)\,df$$

This definition can be extended to bandpass signals and to complex signals,[12] but from the standpoint of physical significance, a better measure of mean "bandwidth" is given by[13]**

$$B = \frac{\left(\int_{-\infty}^{\infty} W_E(f)\,df\right)^2}{\int_{-\infty}^{\infty} W_E^2(f)\,df}$$

In this Memorandum, however, the bandwidth of principal interest is one which describes the total or "peak" spectral extent of the power spectrum. Again different definitions are possible, but the

---

*A low-pass signal is one having spectral components distributed more or less continuously from nearly zero frequency to some maximum or cutoff frequency.

**Personal communication with W. Doyle, consultant to The RAND Corporation.

one to be used here gives the (peak) bandwidth of a signal as the (physical) frequency interval that includes all but a specified small fraction β of the total average power of the signal, excluding any dc component.

In the case of the baseband signal $E_B(t)$, with normalized power spectrum $w_{E_B}(f)$, the bandwidth B_B is given by

$$S_B(0,B_B) \Big/ S_B \equiv \int_{-B_B}^{B_B} w_{E_B}(f)\, df = 1 - \beta \tag{64}$$

Note the similarity to the definition of peak signal power given in Eq. (50). In the case of the modulated carrier $E_R(t)$, the bandwidth B_R for a carrier with double sidebands is given by

$$S_R\left(f_o - \frac{1}{2} B_R,\ f_o + \frac{1}{2} B_R\right) \Big/ S_R = 2 \int_{f_o - \frac{1}{2} B_R}^{f_o + \frac{1}{2} B_R} w_{E_R}(f)\, df = 1-\beta \tag{65}$$

The expressions for single-sideband techniques are the same except that the range of integration is f_o to $f_o + B_R$ for an upper sideband, and $f_o - B_R$ to f_o for a lower sideband. In evaluating the integrals in Eqs. (63) and (65), any carrier frequency component in the power spectrum should be ignored since it carries no information and has zero spectral extent.

The value of β used in defining bandwidth should be chosen in somewhat the same fashion as was the probability ϵ used in defining peak power, i.e., to make acceptably low the distortion that results from

rejecting power outside the bandwidth B_R. On this basis, B_R is then an appropriate measure of the passbands to be provided by the RF and IF filters of the system. As such, it is also a measure of the total noise power admitted to the receiver.* In addition B_R determines the RF channel bandwidth that must be allocated to the modulated carrier if adjacent channel interference is to be avoided.

Probability Density Functions--Peak and Average Power

The practical significance of the average and peak values of the modulated carrier power at the receiver input has already been indicated. The straightforward approach to calculating these quantities for a particular modulating method and baseband would be to determine the probability density function $p_{E_R}(x)$ corresponding to the received carrier

$$E_R(t) = A_R\, e(t) = A_R\, a(t)\, \cos\left[\omega_o t + \varphi(t)\right]$$

and then apply the definitions of Eqs. (48) and (50) to obtain S_R and $\hat{S}_R$ respectively.

Unfortunately, although $a(t)$ and $\varphi(t)$ are fairly simple functions of the baseband $e_B(t)$ for most analog modulation methods (see Table 1), the statistical nature of $e_B(t)$ can easily be such as to make $p_{E_R}(x)$ extremely difficult to calculate. In such cases, it is helpful to assume that the amplitude and phase dependent factors in $E_R(t)$ are statistically independent. The average power is then the product of the mean square values of these factors

*But B_R may well differ from the "noise bandwidth" of the receiver, defined as the width of the rectangular filter which has the same center frequency gain and passes the same white noise power as does the receiver.

$$S_R = \overline{E_R^2(t)} = \overline{\left[A_R\, a(t)\right]^2}\ \overline{\cos^2\left[\omega_o t + \varphi(t)\right]}$$

If it is also assumed that the phase is uniformly distributed, then

$$\overline{\cos^2\left[\omega_o t + \varphi(t)\right]} = \tfrac{1}{2}$$

and so

$$S_R = \overline{a^2(t)}\ S_{Ro}$$

where

$$S_{Ro} = A_R^2/2$$

is the average power of the unmodulated carrier. In this way, calculation of the average carrier power is reduced to determining the mean square value of the amplitude ratio

$$S_R/S_{Ro} = \overline{a^2(t)} = \int_{-\infty}^{\infty} x^2 p_a(x)dx \qquad (66)$$

This is a much simpler task since only the probability density $p_a(x)$ of the amplitude ratio is needed, and, in most cases, it can be expressed directly in terms of the probability density $p_{e_B}(x)$ of the modulating baseband.

Calculation of the peak power $\hat{S}_R$ can also be made to depend only on $p_a(x)$ if it is noted that the excursions of $E_R(t)$ are bounded by the envelope $|A_R\, a(t)|$ and that, for the narrow-band signals here assumed, this envelope varies quite slowly compared with $E_R(t)$ itself. It is apparent that, whatever the value of ϵ used in Eq. (50) to define the peak power $\hat{S}_R$ corresponding to a particular modulated carrier, there is an α such that if

$$\Pr\left\{ A_R^2 |a(t)|^2 > \hat{S}_R \right\} = \alpha \tag{67}$$

then

$$\Pr\left\{ E_R^2(t) > \hat{S}_R \right\} = \epsilon$$

The appropriate value of α is of course determined by distortion measurements similar to those required in the specification of ϵ. In terms of $p_a(x)$, the condition expressed by Eq. (67) may be written

$$\left.\begin{aligned} &\int_{-\hat{a}}^{\hat{a}} p_a(x)dx = 1-\alpha \\ &\text{where} \\ &\hat{a} \equiv \frac{\sqrt{\hat{S}_R}}{A_R} \equiv \sqrt{\frac{\hat{S}_R}{2S_{Ro}}} \end{aligned}\right\} \tag{68}$$

This definition of peak carrier power is equivalent to, and henceforth will be used in place of, Eq. (50). The value of $\hat{S}_R$ so determined is closely related to the so-called "peak envelope power" S_{PEP}, defined loosely as the mean power at the crest of the modulation envelope. Thus

$$S_{PEP} = \frac{1}{2} \hat{S}_R$$

In the case of the baseband signal $E_B(t)$, the peak power $\hat{S}_B$ is defined by the condition

$$\int_{-\sqrt{\hat{S}_B}}^{\sqrt{\hat{S}_B}} p_{E_B}(x)\, dx = 1-\alpha \tag{69}$$

where $p_{E_B}(x)$ is the baseband probability density function and α has the same value used in Eq. (68) to define the peak received carrier power. The normalized baseband is then

$$e_B(t) = \frac{E_B(t)}{\sqrt{\hat{S}_B}}$$

and the baseband peak-to-average power ratio becomes just the reciprocal of its mean square value:

$$\Lambda_B \equiv \frac{\hat{S}_B}{S_B} = \frac{1}{\overline{e_B^2(t)}} \tag{70}$$

Arbitrary Basebands. Equations (66), (68), and (70) can now be used in conjunction with the values of a(t) given in Table 1 to obtain general expressions for S_R, $\hat{S}_R$, and $\Lambda_R = \hat{S}_R/S_R$ in terms of parameters which describe the modulation method and the baseband. Table 3 displays such expressions for each of the modulation methods considered in this report.*

Referring first to the entries for the double-sideband methods (AM, DSB, PM, FM), it is seen that S_R, $\hat{S}_R$, and Λ_R depend only on the modulation index and the baseband peak-to-average power ratio Λ_B. Indeed, for the angle modulation techniques, modulation has no effect at all on the carrier power.

In the case of AM, note that the average carrier power is made up of two terms. The first is of course the average carrier power in the absence of modulation, while the second represents the total power

*Table 3 also includes the expressions for RF bandwidth B_R derived later in this section.

Table 3

RF POWER AND BANDWIDTH OF A SINUSOIDAL CARRIER,
ANALOG MODULATED BY AN ARBITRARY BASEBAND

Modulation Method	Amplitude Ratio $a(t)$	Normalized Average Received Carrier Power S_R/S_{Ro}	Normalized Peak Received Carrier Power $\hat{S}_R/S_{Ro}$	Peak-to-Average Carrier Power Ratio Λ_R	RF Bandwidth Ratio B_R/B_B
AM	$1 + Me_B(t)$	$1 + M^2/\Lambda_B$	$2(1 + M)^2$	$\frac{(1 + M)^2}{\Lambda_B + M^2} 2\Lambda_B$	2
DSB	$Me_B(t)$	M^2/Λ_B	$2M^2$	$2\Lambda_B$	2
SSB	$M\sqrt{e_B^2(t) + \breve{e}_B^2(t)}$	$2M^2/\Lambda_B$	$2\hat{a}^2$	$\frac{\hat{a}^2}{M^2}\Lambda_B$	1
PM	1	1	2	2	$\sim 2\left(1 + \frac{\Phi}{\sqrt{3}}\right)$
FM	1	1	2	2	$\sim 2(1 + D)$
SSBFM	$e^{\mp\Omega \int \breve{e}_B(t)\, dt}$	$\overline{a^2(t)}$	$2\hat{a}^2$	$\frac{2\hat{a}^2}{\overline{a^2(t)}}$	$\sim \frac{4}{3}(1 + D)$

carried by the two sidebands. The sideband power is thus smaller than the unmodulated carrier power by the factor M^2/Λ_B. It is only in the special case of 100 percent modulation (M = 1) by a sinusoidal baseband ($\Lambda_B = 2$) that the sideband power equals one-half the unmodulated carrier power.

In the case of DSB, there is no carrier component and all of the RF power is in the sidebands. But relative to the average power that would be received if only the unmodulated carrier were transmitted, the average sideband power is the same as for an AM signal with the same modulation index. In order to radiate the same average sideband power as a DSB transmitter, an AM transmitter with modulation index M must therefore radiate a total average power that is higher by the factor $1 + \Lambda_B/M^2$. The AM transmitter must also have a higher peak power capability. For the same average sideband power, Table 3 shows that the peak carrier power with AM is $(1 + M)^2/M^2$ times higher than with DSB. For example, if the AM transmitter is 100 percent modulated (M = 1), it must have four times the peak power capability.*

The situation is more complicated in the case of the linear and exponential single-sideband techniques. Here, the amplitude ratio of the modulated carrier depends on both the baseband and its Hilbert transform. And although these two waveforms have equal variances and so represent the same average power, they are uncorrelated at a given

*Although both average and peak power are higher with AM, it should be noted that the RF peak-to-average power ratio for AM differs from that for DSB by the factor $(1 + M)^2/(\Lambda_B + M^2)$. For example, with 100 percent modulation (M = 1) and an FDM speech baseband ($\Lambda_B \cong 10$), Λ_R is equal to 20 (13 db) for DSB, and only 7.3 (8.6 db) for AM.

instant and in general do not have the same probability density functions.[9] For this reason, the general expressions for the peak and average powers of the single-sideband modulated carriers have, with one exception, been left in terms of the peak and average values of the normalized instantaneous envelope power $a^2(t)$. The exception is the average power of the SSB carrier. By virtue of the equality between the variances of $e_B(t)$ and $\breve{e}_B(t)$, Eqs. (66) and (70) give

$$S_R/S_{Ro} = M^2\left[\overline{e_B^2(t)} + \overline{\breve{e}_B^2(t)}\right] = 2M^2\overline{e_B^2(t)} = 2M^2/\Lambda_B \qquad \text{SSB}$$

The peak SSB power and both the average and the peak SSBFM power can be calculated only in specific cases where the statistics of the Hilbert transformed baseband are known.

Sinusoidal and Gaussian Basebands. The general expressions for average and peak received carrier power given in Table 3 will be illustrated by considering two representative types of baseband:* a sinusoidal test tone

$$E_B(t) = a\cos\omega_1 t$$

and a random signal specified only by the condition that its probability density function is gaussian with zero mean and variance $\sigma_{E_B} = \overline{E_B^2(t)}$. The probability density functions for these basebands are respectively

$$p_{E_B}(x) = \begin{cases} \left(\pi\sqrt{a^2 - x^2}\right)^{-1} & |x| < a \\ 0 & |x| > a \end{cases} \qquad \text{test tone} \quad (71)$$

*Multichannel speech signals will be considered in the next section.

and

$$p_{E_B}(x) = \left(\sqrt{2\pi}\,\sigma_{E_B}\right)^{-1} \exp\left(-x^2/2\sigma_{E_B}^2\right) \qquad \text{gaussian} \qquad (72)$$

From these, the average and peak baseband powers may be computed using Eqs. (48) and (69) respectively. The resultant expressions may then be combined to yield the dependence of the baseband peak-to-average power ratio Λ_B on the probability α that $E_B(t)$ exceeds $\sqrt{\hat{S}_B}$

$$\Lambda_B = 2\cos^2(\alpha\pi/2) \qquad \text{test tone} \qquad (73)$$

$$\operatorname{erfc}(\sqrt{\Lambda_B/2}) = \alpha \qquad \text{gaussian} \qquad (74)$$

where

$$\operatorname{erfc}(x) = 1 - \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-y^2}\,dy \qquad (75)$$

is the complementary error function. Plots of Λ_B versus α for these two basebands are shown in Fig. 3. Note from Fig. 3b that with a gaussian baseband, the value of Λ_B is quite sensitive to the choice of α, whereas with a test tone, $\Lambda_B \simeq 2$ for all small values of α. It follows that, with the sinusoidal test tone, the normalized baseband waveform is $e_B(t) = \cos \omega_1 t$ for all cases of practical interest.

As shown by the entries in Table 3, these data on the baseband peak-to-average power ratio are sufficient to determine S_R, $\hat{S}_R$, and Λ_R for test-tone and gaussian modulated carriers in the case of all modulation methods except SSB and SSBFM. For SSB, it is first necessary to compute the peak value of the instantaneous normalized envelope

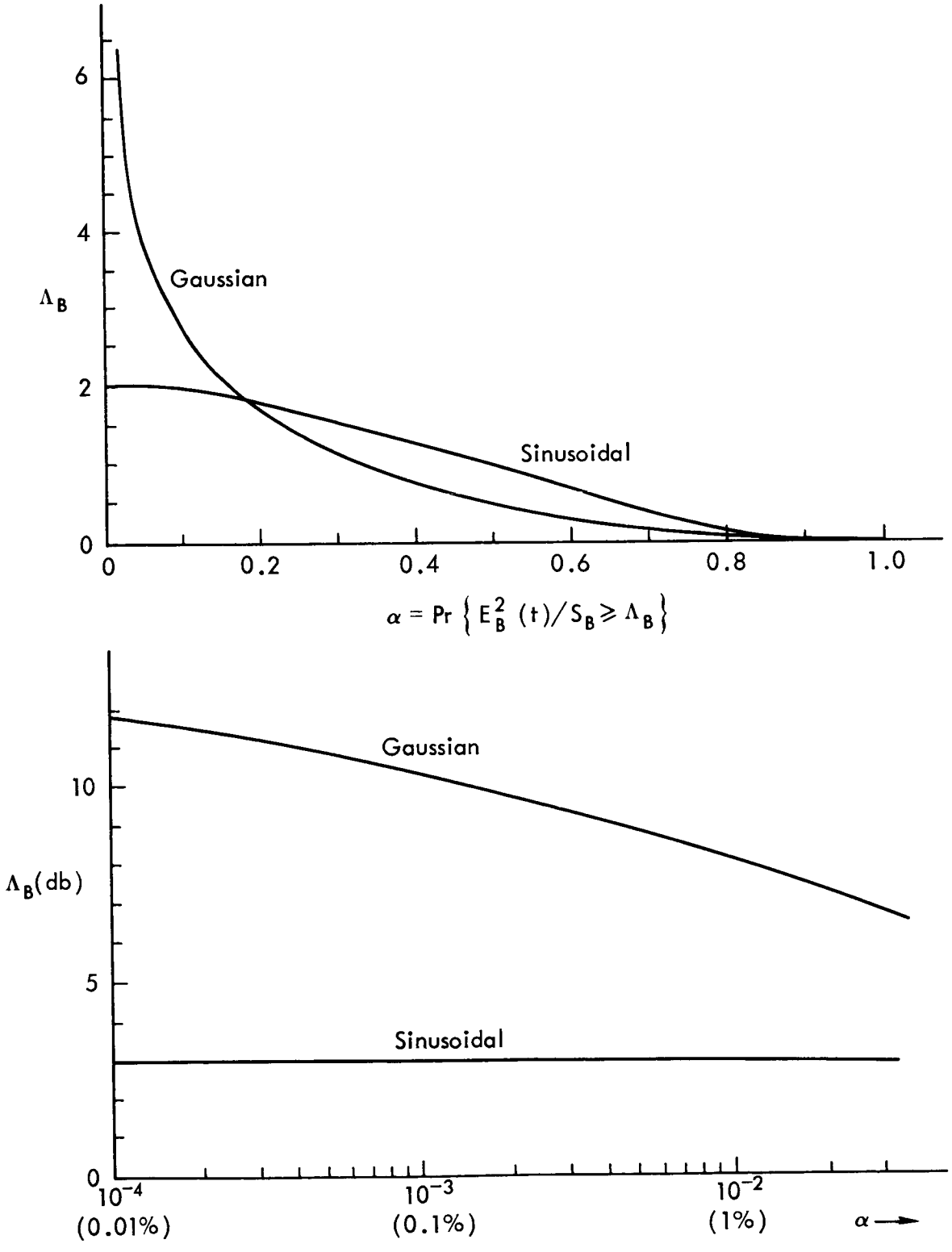


Fig. 3—Peak-to-average power ratio Λ_B versus overload probability α for sinusoidal and gaussian basebands

$$a(t) = M\sqrt{e_B^2(t) + \breve{e}_B^2(t)}$$

In the case of the test tone $e_B(t) = \cos \omega_1 t$, Eq. (15) gives $e_B(t) = \sin \omega_1 t$. Therefore, $a(t) = M$, $\hat{a} = M$, and from Table 3,

$$\left.\begin{aligned} \hat{S}_R/S_{Ro} &= 2\hat{a}^2 = 2M^2 \\ \Lambda_R &= \frac{\hat{a}^2}{M^2}\Lambda_B = \Lambda_B = 2 \end{aligned}\right\} \quad \text{SSB by test tone} \qquad (76)$$

In contrast, note from Table 3 that $\Lambda_R = 2\Lambda_B = 4$ for DSB modulation by a test tone. In the case of a gaussian $e_B(t)$, it is easily shown that $\breve{e}_B(t)$ is also gaussian and, as a result, $a(t)$ is Rayleigh distributed[11] with probability density

$$p_a(x) = \begin{cases} \dfrac{x}{\sigma_{e_B}^2} \exp\left(-x^2/2\sigma_{e_B}^2\right) & x \geq 0 \\ 0 & x < 0 \end{cases} \qquad (77)$$

Substituting this into Eq. (66) and (68) to compute S_R/S_{Ro} and $\hat{S}/S_{Ro}$ respectively, and combining the results to eliminate σ_{e_B}, the RF peak-to-average ratio can be expressed directly in terms of α:

$$\Lambda_R = \ln(\alpha^{-2}) \qquad \text{SSB by gaussian} \qquad (78)$$

With SSBFM, the general expression for the amplitude ratio is

$$a(t) = \exp\left(\mp \Omega \int \breve{e}_B(t)dt\right)$$

In the case of the sinusoidal test tone, $e_B(t) = \cos \omega_1 t$, this becomes

$$a(t) = \exp(\pm D \cos \omega_1 t) \tag{79}$$

where

$$D \equiv \Omega/\omega_1$$

is the deviation ratio or modulation index for a single sinusoid of angular frequency ω_1. The corresponding probability density function is

$$p_a(x) = \begin{cases} \left(\pi x \sqrt{D^2 - \ln^2 x}\right)^{-1}, & e^{-D} < x < e^{D} \\ 0, & \text{otherwise} \end{cases} \tag{80}$$

Substituting either Eq. (79) or Eq. (80) into Eq. (66), the normalized average RF power is given by

$$S_R/S_{Ro} = I_o(2D)$$

where $I_o(x) = J_o(jx)$ is the real-valued even function of x given by the zero order Bessel function with a purely imaginary argument. Substituting Eq. (80) into Eq. (68), the normalized peak RF power is

$$\hat{S}_R/S_{Ro} = 2 \exp(2D \cos \alpha\pi) \cong 2 \exp(2D) \quad , \quad \alpha \ll 1$$

from which

$$\Lambda_R = 2 \exp(2D)/I_o(2D) \qquad \text{SSBFM by test tone} \tag{81}$$

In the case of the gaussian baseband with probability density function given by Eq. (72), the probability density function for the amplitude ratio of the SSBFM modulated carrier is

$$p_a(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\ln^2 x/2\sigma^2}, & x > 0 \\ 0, & x < 0 \end{cases} \tag{82}$$

where σ^2 is the variance of the normally distributed baseband-dependent integral $\Omega \int \check{e}_B(t)dt$. To express σ^2 in terms of the baseband parameters requires a knowledge of the second-order statistics of the baseband, as given by its power spectrum. If this is assumed to be flat over the bandwidth of the baseband B_B, and zero elsewhere, it can easily be shown that

$$\sigma^2 = \Omega^2 \int_{-\infty}^{\infty} \omega^{-2} W_{e_B}(f)df = \frac{f_m}{f_\ell} \frac{D^2}{\Lambda_B} \tag{83}$$

where f_ℓ and $f_m = f_\ell + B_B$ are respectively the lowest and highest frequencies in the baseband, Λ_B is the baseband peak-to-average ratio given by Eq. (74), and

$$D = \Omega / \omega_m$$

is the FM modulation index for a baseband whose highest angular frequency is $\omega_m = 2\pi f_m$. Substituting Eq. (82) into Eqs. (66) and (68) leads to the result

$$\left.\begin{aligned} \overline{a^2(t)} &= \exp(2\sigma^2) \\ \operatorname{erfc}(\ln \hat{a}/\sqrt{2}\,\sigma) &= 2\alpha \end{aligned}\right\} \quad \text{SSBFM by gaussian}$$

If the inverse complementary error function is denoted by erfc^{-1}, these expressions may be combined to yield

$$\Lambda_R = \exp\{2\sigma[\sqrt{2}\,\operatorname{erfc}^{-1}(2\alpha) - \sigma]\} \qquad \text{SSBFM by gaussian} \qquad (84)$$

where σ depends on α and the baseband and modulation parameters as shown in Eq. (83). It is interesting to note from Eq. (84) that regardless of how small a value of α is selected, a modulation index can be chosen to make $\sigma > \sqrt{2}\,\operatorname{erfc}^{-1}(2\alpha)$ and so lead to a peak-to-average power ratio less than unity.*

To facilitate comparison, the expressions just developed for the peak-to-average carrier power corresponding to various types of analog modulation by sinusoidal and gaussian basebands have been assembled in Table 4. Plots of Λ_R versus α based on these expressions are shown in Fig. 4.

Power Spectra and Bandwidths

In order to calculate the power spectrum $W_{E_R}(f)$ of the modulated carrier, and from it calculate the RF bandwidth B_R defined in Eq. (65), it is useful to have expressions relating $W_{E_R}(f)$ to the power spectra of the complex modulated carrier $c(t)$ and the modulation function $m(t)$ introduced in Eq. (34). The waveforms in question bear the following relationships to one another

*Note, however, that the resultant wide dynamic range of the modulated carrier makes it impractical to use such large modulation indices with SSBFM.

Table 4

PEAK-TO-AVERAGE RF POWER RATIOS Λ_R FOR CARRIERS MODULATED BY SINUSOIDAL AND GAUSSIAN BASEBANDS

Modulation Method	Sinusoidal	Gaussian
AM	$\frac{(1+M)^2}{\Lambda_B+M^2} 4 \cos^2\left(\alpha \frac{\pi}{2}\right)$ $\cong 4 \frac{(1+M)^2}{\Lambda_B+M^2}$	$4 \frac{(1+M)^2}{\Lambda_B+M^2} (\text{erfc}^{-1} \alpha)^2$
DSB	$4 \cos^2\left(\alpha \frac{\pi}{2}\right)$ $\cong 4$	$4(\text{erfc}^{-1} \alpha)^2$
SSB	$2 \cos^2\left(\alpha \frac{\pi}{2}\right)$ $\cong 2$	$\ln (\alpha^{-2})$
PM	$2 \cos^2\left(\alpha \frac{\pi}{2}\right)$ $\cong 2$	$2 \cos^2\left(\alpha \frac{\pi}{2}\right)$ $\cong 2$
FM	$2 \cos^2\left(\alpha \frac{\pi}{2}\right)$ $\cong 2$	$2 \cos^2\left(\alpha \frac{\pi}{2}\right)$ $\cong 2$
SSBFM	$\frac{2 \exp(2D \cos \alpha\pi)}{I_o(2D)}$ $\cong 2e^{2D}/I_o(2D)$	$\exp\left\{2\sigma\left[\sqrt{2}\ \text{erfc}^{-1}(2\alpha)-\sigma\right]\right\}$ where $\sigma = \sigma(\alpha) = \frac{f_m}{f_\ell}\left(\frac{D}{2\ \text{erfc}^{-1}\ \alpha}\right)^2$

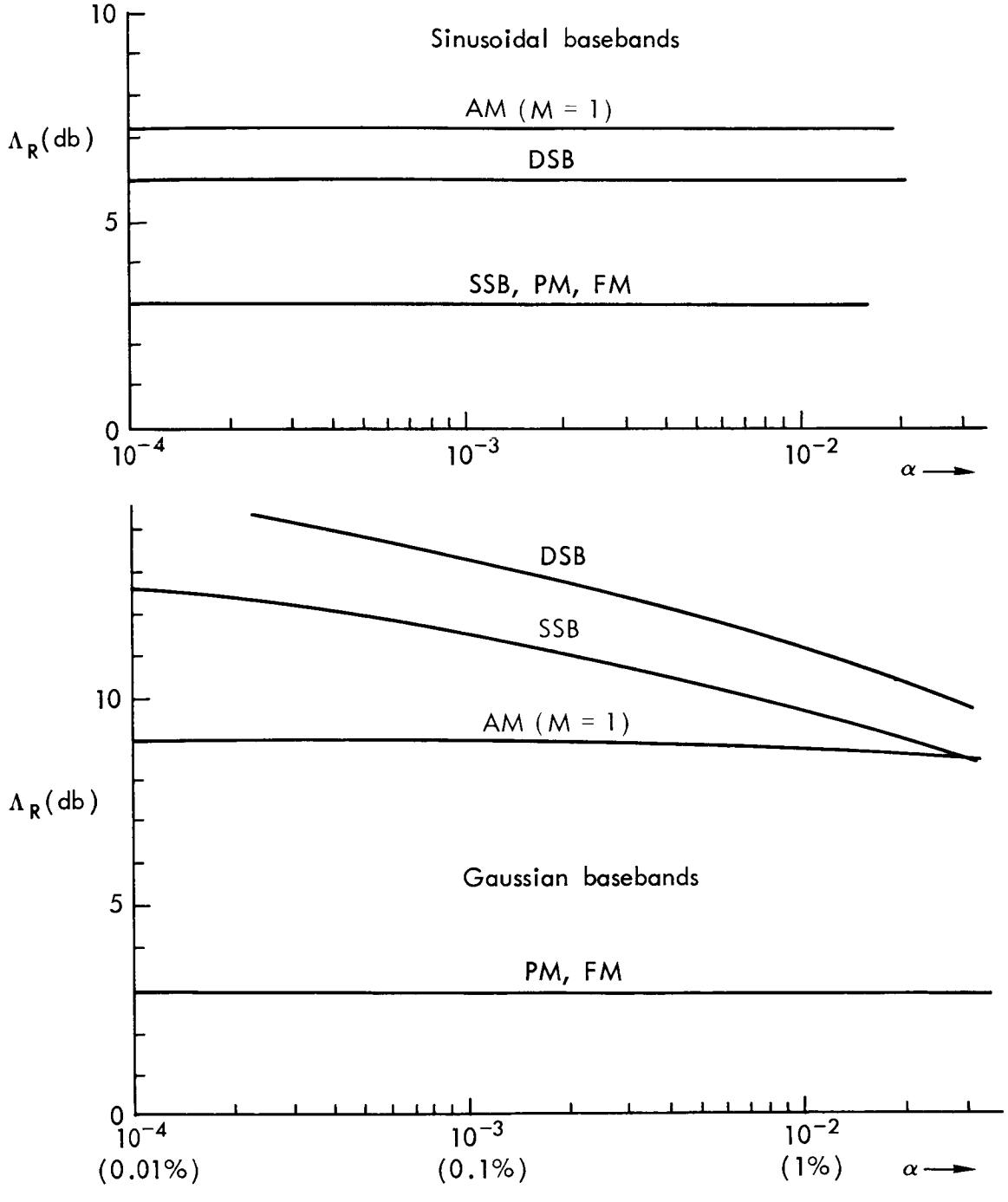


Fig. 4—Peak-to-average RF power ratio Λ_R versus overload probability α for carriers modulated by sinusoidal and gaussian basebands

$$E_R(t)/A_R = e(t) = \mathrm{Re}\ c(t) \tag{85}$$

$$c(t) = m(t)\ e^{j\omega_o t} \tag{86}$$

Moreover, for the narrow-band modulated carriers of interest here, $c(t)$ is the analytic signal representation of $e(t)$

$$c(t) = \underline{e}(t) = e(t) + j\ \check{e}(t) \tag{87}$$

The relationships between power spectra may be obtained by first taking the autocorrelation functions of these equations and then applying the Wiener-Khinchin theorem of Eq. (57). For example, using the definition of Eq. (56) on Eq. (85) gives $R_{E_R}(\tau) = A_R^2\, R_e(\tau)$. Taking Fourier transforms and applying Eq. (57) then gives the almost trivial relation between the power spectra of $E_R(t)$ and $e(t)$

$$W_{E_R}(f) = A_R^2\, W_e(f) \tag{88}$$

Applying the same procedure to Eq. (87) gives as a first step

$$R_c(\tau) = R_e(\tau) + R_{\check{e}}(\tau) - j\left[R_{\check{e}e}(\tau) - R_{e\check{e}}(\tau)\right] \tag{89}$$

where notation of the form $R_{xy}(\tau)$ represents the "cross-correlation function"

$$R_{xy}(\tau) = \overline{x^*(t)\ y(t+\tau)} \tag{89a}$$

This may be simplified by noting[9] that an arbitrary real waveform $x(t)$ and its Hilbert transform have the same autocorrelation function

$$R_x(\tau) = R_{\breve{x}}(\tau) \tag{89b}$$

Also, the cross-correlations of $x(t)$ and $\breve{x}(t)$ are given by the Hilbert transform of $R_x(t)$

$$R_{\breve{x}x}(\tau) = R_{x\breve{x}}(-\tau) = - R_{x\breve{x}}(\tau) = - \breve{R}_x(\tau) \tag{89c}$$

and from this, it follows that $x(t)$ and $\breve{x}(t)$ are uncorrelated at the same instant:

$$R_{\breve{x}x}(0) = R_{x\breve{x}}(0) = \breve{R}_x(0) = 0 \tag{89d}$$

Applying these to Eq. (89) yields

$$R_c(\tau) = 2\left[R_e(\tau) + j\, \breve{R}_e(\tau)\right] = 2\, \underset{\sim}{R}_e(\tau) \tag{90}$$

which shows that, just as $c(t)$ is the analytic signal representation of $e(t)$, the autocorrelation of $c(t)$ is the analytic signal representation of twice the autocorrelation of $e(t)$. Hence

$$R_e(\tau) = \frac{1}{2}\, \mathrm{Re}\left[R_c(\tau)\right] = \frac{1}{4}\left[R_c(\tau) + R_c^*(\tau)\right]$$

and since by Eq. (60), $R_c^*(\tau) = R_c(-\tau)$, application of the Wiener-Khinchin theorem yields the desired relation between the power spectra of $e(t)$ and $c(t)$

$$W_e(f) = \frac{1}{4}\left[W_c(f) + W_c(-f)\right] \tag{91}$$

In this relation it should be noted that since $W_c(f)$ is the Fourier spectrum of the analytic signal $R_c(\tau)$, it vanishes for negative

frequencies. Similarly, $W_c(-f)$ vanishes for positive frequencies, and the two terms in Eq. (91) share at most a dc component. Thus, the power spectrum of the real modulated carrier can be obtained by adding to the power spectrum of the complex carrier its reflection about zero frequency, and then dividing the magnitude of all components by 4.

The relation between the power spectra of c(t) and m(t) may be found in similar fashion. Applying Eq. (56) to Eq. (86) yields the relation between the autocorrelation functions:

$$R_c(\tau) = R_m(\tau)\, e^{j2\pi f_o \tau}$$

and application of the Wiener-Khinchin theorem then gives

$$W_c(f) = W_m(f - f_o) \tag{92}$$

This shows that the power spectrum of the complex modulated carrier can be obtained by shifting the power spectrum of the modulation function upward in frequency by an amount equal to the carrier frequency. Finally, the relation between the power spectra of e(t) and m(t) is obtained by combining Eq. (91) and (92) to give

$$W_e(f) = \tfrac{1}{4} \left\{ W_m(f - f_o) + W_m\left[- (f + f_o) \right] \right\} \tag{93a}$$

or, if m(t) is real,

$$W_e(f) = \tfrac{1}{4} \left[W_m(f - f_o) + W_m(f + f_o) \right] \tag{93b}$$

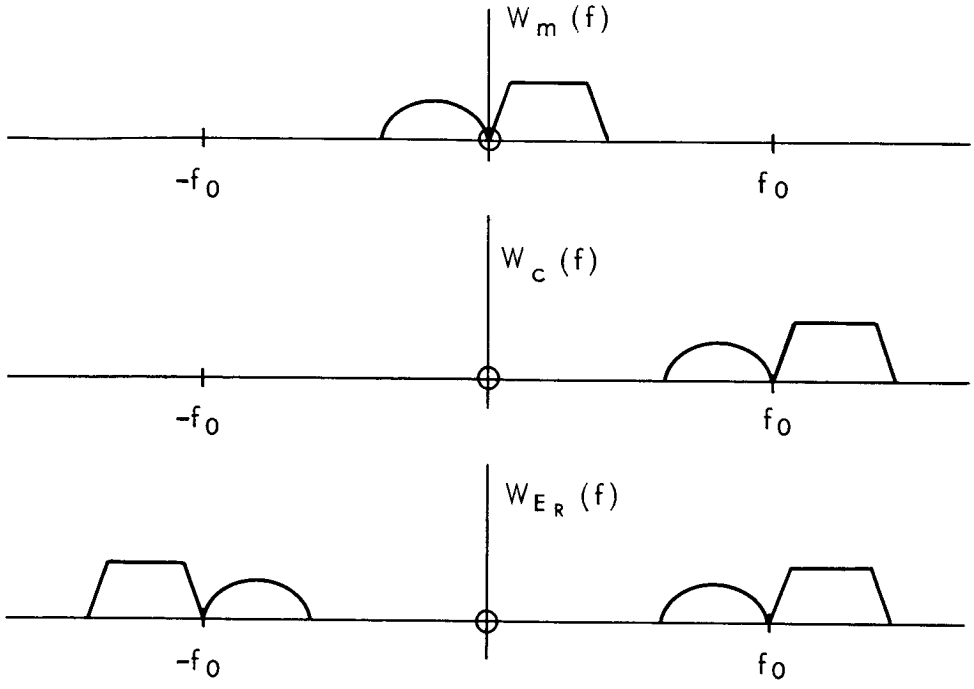
since in this case, $W_m(f)$ is an even function of frequency. Note in each of these equations that, just as in Eq. (91), the first term vanishes for negative frequencies and the second for positive frequencies.

A comparison of Eqs. (91), (92), and (93) with Eqs. (41), (35), and (42) respectively shows the close similarity between the relations for the power spectra and those for the Fourier spectra. More importantly, Eq. (93) reduces the task of computing the power spectrum of the normalized real modulated carrier e(t) to that of finding the spectrum of the modulation function m(t). Combining this with Eq. (88) it follows that $W_{E_R}(f)$ can be obtained by shifting $W_m(f)$ upward in frequency by an amount equal to the carrier frequency, adding its negative-frequency reflection, and then multiplying all spectral components by $A_R^2/4$.

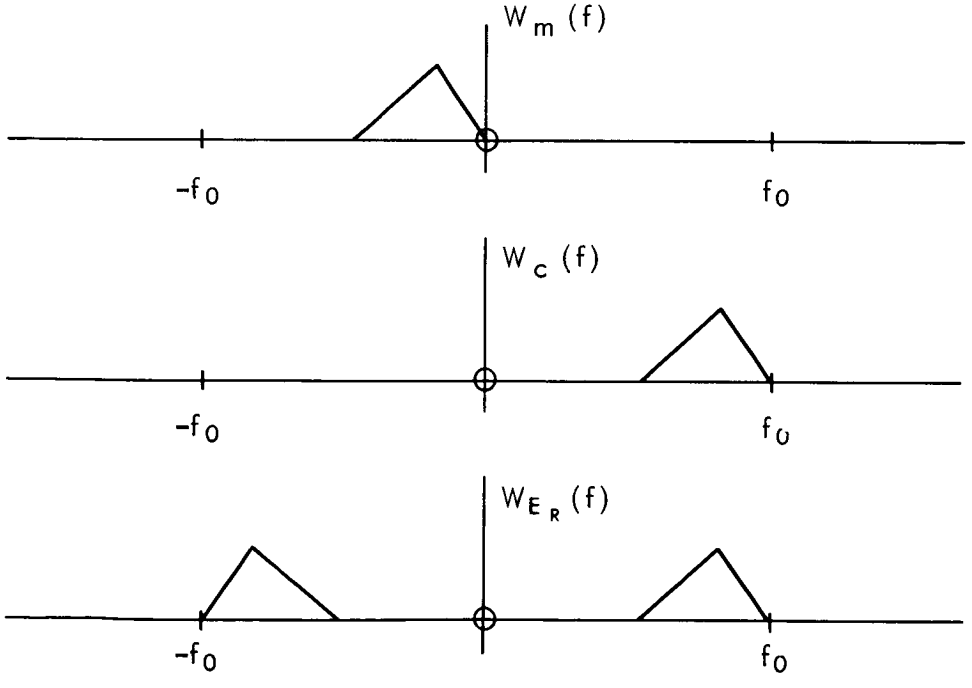
The relationships just described between the power spectra of c(t), m(t) and $E_R(t)$ are depicted schematically* in Fig. 5. From these relations it is apparent that the bandwidth B_R of the real modulated carrier, defined as in Eq. (65) by the range of physical frequencies required to include all but a fraction β of the total carrier power, is identical to the similarly defined bandwidth of the complex modulated carrier c(t). For double-sideband modulation methods, B_R is then exactly twice the bandwidth occupied by the modulation function m(t), whereas for single-sideband methods, B_R is just equal to the bandwidth of m(t).

---

*The arbitrary shapes of the sidebands in Fig. 5a were chosen to illustrate general relationships. In practice, upper and lower sidebands are most often symmetrical about the carrier frequency.



a. Double-sideband modulation methods



b. Single-sideband modulation methods
(Lower sideband illustrated)

Fig.5—General relations among power spectra

For a formal demonstration of these conclusions, it is only necessary to substitute Eqs. (91) and (93) into Eq. (65). For example, combining Eqs. (93) and (65) yields the following expression for B_R in terms of $W_m(f)$ when m(t) is a double-sideband modulation function

$$\int_{-\frac{1}{2}B_R}^{\frac{1}{2}B_R} W_m(f)\, df = (1 - \beta) \int_{-\infty}^{\infty} W_m(f)\, df \qquad (94)$$

When m(t) represents a single-sideband method the same equation applies, except that the limits of integration on the left-hand side are 0 and B_R for an upper sideband (USB), and $-B_R$ and 0 for a lower sideband (LSB). In using Eq. (94), dc components in $W_m(f)$ should be neglected for reasons similar to those noted in connection with Eq. (65).

Arbitrary Basebands. As just shown, calculation of the power spectrum and bandwidth of the modulated carrier reduces to finding $W_m(f)$ from the statistics of the baseband signal $e_B(t)$ and the functional relation between $e_B(t)$ and m(t). This will be done for each of the analog modulation methods in turn; the resultant expressions for B_R as a function of basebandwidth B_B of the baseband are collected in the last column of Table 3.

In terms of the dependence of m(t) on $e_B(t)$ shown in Table 2, it is seen that DSB affords the simplest example since, for this modulation method, $m(t) = Me_B(t)$. Hence, the power spectra of m(t) and $e_B(t)$ differ only by the factor M^2,

$$W_m(f) = M^2 W_{e_B}(f) \qquad \text{DSB} \qquad (95)$$

and it follows that, since B_R is twice the bandwidth of m(t), the RF bandwidth of a DSB modulated carrier is twice the baseband bandwidth B_B:

$$B_R = 2B_B \qquad \text{DSB}$$

On a purely formal basis, this relation may also be obtained by substituting Eq. (95) into Eq. (94) and comparing the result with Eq. (64), the definition of B_B.

In the case of AM, $m(t) = 1 + Me_B(t)$, and the relation between power spectra is found by first relating autocorrelation functions. Thus,

$$R_m(\tau) = \overline{\left[1 + Me_B(t)\right]\left[1 + Me_B(t + \tau)\right]} = 1 + M^2 R_{e_B}(\tau)$$

and applying the Wiener-Khinchin relation

$$W_m(f) = \delta(f) + M^2 W_{e_B}(f) \qquad \text{AM} \qquad (96)$$

where $\delta(f) = \mathfrak{F}(1)$ is the Dirac delta function defined in Eq. (55). Thus, with AM, the modulation function power spectrum differs from the baseband spectrum only by a constant factor and a dc component. It follows that the modulation function bandwidth, and with it, the modulated carrier bandwidth, is again just twice the bandwidth B_B of the baseband:

$$B_R = 2B_B \qquad \text{AM}$$

As before, this result can be obtained formally, here by substituting Eq. (96) into Eq. (94), ignoring the delta function contribution, and comparing the result with Eq. (64) for B_B.

For SSB, Table 2 shows

$$m(t) = M\left[e_B(t) \pm j\breve{e}_B(t)\right] \qquad \text{SSB} \qquad (97)$$

where the upper sign leads to the USB and the lower sign to the LSB. Comparing Eq. (97) with Eq. (87), it is seen that in this case the modulation function $m(t)$ bears the same relation to the baseband $e_B(t)$ that the complex carrier $c(t)$ bears to the real carrier $e(t)$. Thus, by analogy with Eq. (90), it follows that

$$R_m(\tau) = 2M^2\left[R_{e_B}(\tau) \pm j\,\breve{R}_{e_B}(\tau)\right]$$

and hence, by Eq. (29), with $\underline{x}(t)$ replaced by $R_m(\tau)$, and $x(t)$ by $2M^2\,R_{e_B}(\tau)$

$$W_m(f) = \begin{cases} \left[1 + \text{sgn}(f)\right] 2M^2\, W_{e_B}(f), & \text{USB} \\ \left[1 - \text{sgn}(f)\right] 2M^2\, W_{e_B}(f), & \text{LSB} \end{cases} \qquad (98)$$

Thus, $W_m(f)$ can be obtained from $W_{e_B}(f)$ by merely suppressing either its negative or positive frequency components and multiplying the rest

by $4M^2$. As a consequence, the bandwidth of the modulation function is equal to B_B and since, for all single-sideband modulation methods, the former bandwidth is equal to B_R, the well-known conclusion is that for SSB,

$$B_R = B_B \qquad \text{SSB}$$

In this case, a formal derivation involves the substitution of Eq. (98) into Eq. (94) with the appropriate integration limits, followed by comparison with Eq. (64).

For the linear modulation methods just considered, it has been easy to express the power spectrum and the peak bandwidth of the modulated carrier directly in terms of the corresponding baseband quantities, regardless of the nature of the baseband. The situation can be summed up by stating that linear modulation merely translates the baseband spectrum unchanged to the vicinity of the carrier frequency. A carrier component may be added or a sideband subtracted, but what remains is a faithful replica of the baseband spectrum. This simple state of affairs is of course a direct result of the linear relation between the modulation function m(t), which characterizes the modulated carrier, and the baseband function b(t), which is a simple function of the baseband signal. For the exponential modulation methods, this linear relation is replaced, as the name implies, by an exponential one, and the relation between the power spectra of the modulated carrier and the baseband does depend on the statistical nature of the latter.

In the case of the double-sideband exponential techniques, for example, Table 2 gives

$$m(t) = e^{jb(t)} \qquad \text{PM, FM}$$

where $b(t)$ may be identified as the instantaneous phase deviation $\varphi(t)$. Proceeding as before, the first step in finding the power spectrum of $m(t)$ is to calculate its autocorrelation function

$$R_m(\tau) = \overline{m^*(t)\, m(t+\tau)} = \overline{e^{-j[b(t) - b(t+\tau)]}} \tag{99}$$

Regarded as an ensemble average, straightforward evaluation of this expression would require the bivariate probability density function, $p_{b_1 b_2}(x_1, x_2)$, for the random variables

$$b_1 = b(t)$$

$$b_2 = b(t + \tau)$$

However, this probability density can equally well be specified by giving its two-dimensional Fourier transform, the characteristic function,[6]

$$C_{b_1 b_2}(\xi_1, \xi_2) = \int_{-\infty}^{\infty} dx_1\, e^{j\xi_1 x_1} \int_{-\infty}^{\infty} dx_2\, e^{j\xi_2 x_2}\, p_{b_1 b_2}(x_1, x_2) \tag{100}$$

$$= \overline{e^{j(\xi_1 b_1 + \xi_2 b_2)}}$$

Comparing the right-hand side of Eq. (99) with that of Eq. (100), it is seen that for the double-sideband exponential modulation methods, the autocorrelation of the modulation function is given directly by the characteristic function of the second-order probability density of the baseband function evaluated at the point $\xi_1 = -1$, $\xi_2 = 1$:

$$R_m(\tau) = C_{b_1 b_2}(-1, 1) \quad \text{PM, FM} \tag{101}$$

The desired power spectrum $W_m(f)$ is the Fourier transform of this expression and therefore also depends on the second-order statistics of $b(t)$ and hence of the baseband.

A similar situation prevails with the single-sideband exponential methods, except that here the second-order statistics of $\breve{b}(t)$, the Hilbert transform of $b(t)$, are also involved. Thus, from Table 2,

$$m(t) = e^{j[b(t) \pm j\,\breve{b}(t)]} \quad \text{SSBPM, SSBFM} \tag{102}$$

where $b(t)$ is again the instantaneous phase deviation, and the upper and lower signs again correspond respectively to the upper and lower sideband forms of the modulated carrier. Applying Eq. (56), the autocorrelation of $m(t)$ is

$$R_m(\tau) = \overline{e^{j[-b(t) \pm j\,\breve{b}(t) + b(t+\tau) \pm j\,\breve{b}(t+\tau)]}} \tag{103}$$

If the characteristic function of the fourth-order probability density of the random variables

$$b_1 = b(t), \quad b_2 = b(t + \tau), \quad b_3 = \check{b}(t), \quad b_4 = \check{b}(t + \tau)$$

is written

$$C_{b_1 b_2 b_3 b_4}(\xi_1, \xi_2, \xi_3, \xi_4) = e^{\overline{j \sum_{i=1}^{4} \xi_i b_i}} \tag{104}$$

then comparison of Eqs. (103) with (104) gives

$$R_m(\tau) = C_{b_1 b_2 b_3 b_4}(-1, 1, \pm j, \pm j) \qquad \text{SSBPM, SSBFM} \tag{105}$$

Again, to proceed further requires specification of the statistics of $b(t)$ and $\check{b}(t)$.

Sinusoidal and Gaussian Basebands. As with the average and peak power calculations, two types of baseband will be considered for purposes of illustration: a sinusoidal test tone, $\cos \omega_1 t$, and a noise-like signal with gaussian probability density given by Eq. (72).

The power spectrum of the test tone is concentrated at the (physical) frequency, $f_1 = \omega_1/2\pi$, and consists of delta functions at $f = -f_1$ and $f = f_1$

$$W_{e_B}(f) = \tfrac{1}{4}\left[\delta(f - f_1) + \delta(f + f_1)\right] \tag{106}$$

Strictly speaking, the bandwidth B_B of such an idealized baseband signal is zero, since all of the signal power is concentrated at a single (physical) frequency. However, if test tones of any frequency up to a maximum f_m are to be passed, the bandwidth of the baseband circuits must be

$$B_B = f_m \tag{107}$$

The power spectrum of the gaussian signal can be specified arbitrarily so long as its integral over all frequencies yields the same average power as the probability density function. For present purposes, the flat bandpass spectrum introduced in connection with Eq. (83) will be assumed.

$$W_{E_B}(f) = \begin{cases} S_B/2B_B, & f_\ell < |f| < f_m \\ 0, & \text{elsewhere} \end{cases} \tag{108}$$

The bandwidth necessary to just pass such a baseband is obviously

$$B_B = f_m - f_\ell$$

but if the lower frequency limit of the baseband is to be arbitrarily low, Eq. (107) applies.

As just demonstrated, these data on the baseband are sufficient to determine the power spectrum $W_{E_R}(f)$ and bandwidth B_R of the modulated carrier for all of the linear modulation techniques. For the exponential methods on the other hand, it was shown that $W_m(f)$ depends on the

higher-order statistics of the baseband function b(t) for the double-sideband techniques and on both b(t) and its Hilbert transform for the single-sideband exponential techniques. Moreover, the dependence of $W_m(f)$ on the baseband itself is even more complex since the relation between b(t) and $e_B(t)$, though usually linear, involves the parameters of the modulator such as Φ and Ω.

For example, with conventional FM, m(t) is given by $m(t) = e^{j\, b(t)}$, where by Table 2, the instantaneous phase deviation is

$$b(t) = \Omega \int e_B(t)\, dt \tag{109}$$

Thus, even with the simple test tone, $e_B(t) = \cos \omega_1 t$,

$$m(t) = e^{j\, D \sin \omega_1 t} \tag{110}$$

where D is the modulation index or derivation ratio:

$$D = \Omega/\omega_1$$

Since, in this special case, the modulation function is deterministic and periodic, its power spectrum is by Eq. (54)

$$W_m(f) = \sum_{n=-\infty}^{\infty} |M_n|^2\, \delta(f - nf_1)\ , \quad f_1 = \omega_1/2\pi \tag{111}$$

where

$$M_n = \frac{1}{T} \int_{-T/2}^{T/2} m(t)\, e^{-jn\omega_1 t}\, dt \tag{112}$$

is the coefficient of the n^{th} harmonic in the Fourier series expansion

$$m(t) = \sum_{n=-\infty}^{\infty} M_n \, e^{jn\omega_1 t} \tag{113}$$

Substituting Eq. (110) into Eq. (112) yields the well-known result

$$M_n = \frac{1}{T} \int_{-T/2}^{T/2} e^{-j(n\omega_1 t - D \sin \omega_1 t)} \, dt = J_n(D)$$

where J_n is the n^{th} order Bessel function. By Eq. (111) then, the power spectrum of the modulation function is

$$W_m(f) = \sum_{n=-\infty}^{\infty} J_n^2(D) \, \delta(f - nf_1) \qquad \text{FM} \tag{114}$$

a series of equally spaced spectral lines located symmetrically about zero frequency at harmonics of the modulating test tone as shown in Fig. 6a for various values of the FM deviation ratio D.

The relative magnitudes of the sideband components are obviously controlled by D. For all values of D, there is a finite dc component (which comprises the carrier frequency component in the modulated carrier signal) and, in general, all harmonics carry some power. Thus, the spectrum is infinite in extent, but $J_n(D)$ approaches zero for $n > D$ in such a way that the power carried by harmonics with order much greater than the deviation ratio is negligibly small. Substituting Eq. (114) into Eq. (94), and noting that the total spectral power $\int_{-\infty}^{\infty} W_m(f) \, df = \sum_{n=-\infty}^{\infty} J_n^2(D) = 1$, independent of D, it follows that the peak bandwidth of the modulated carrier is the smallest value of B_R

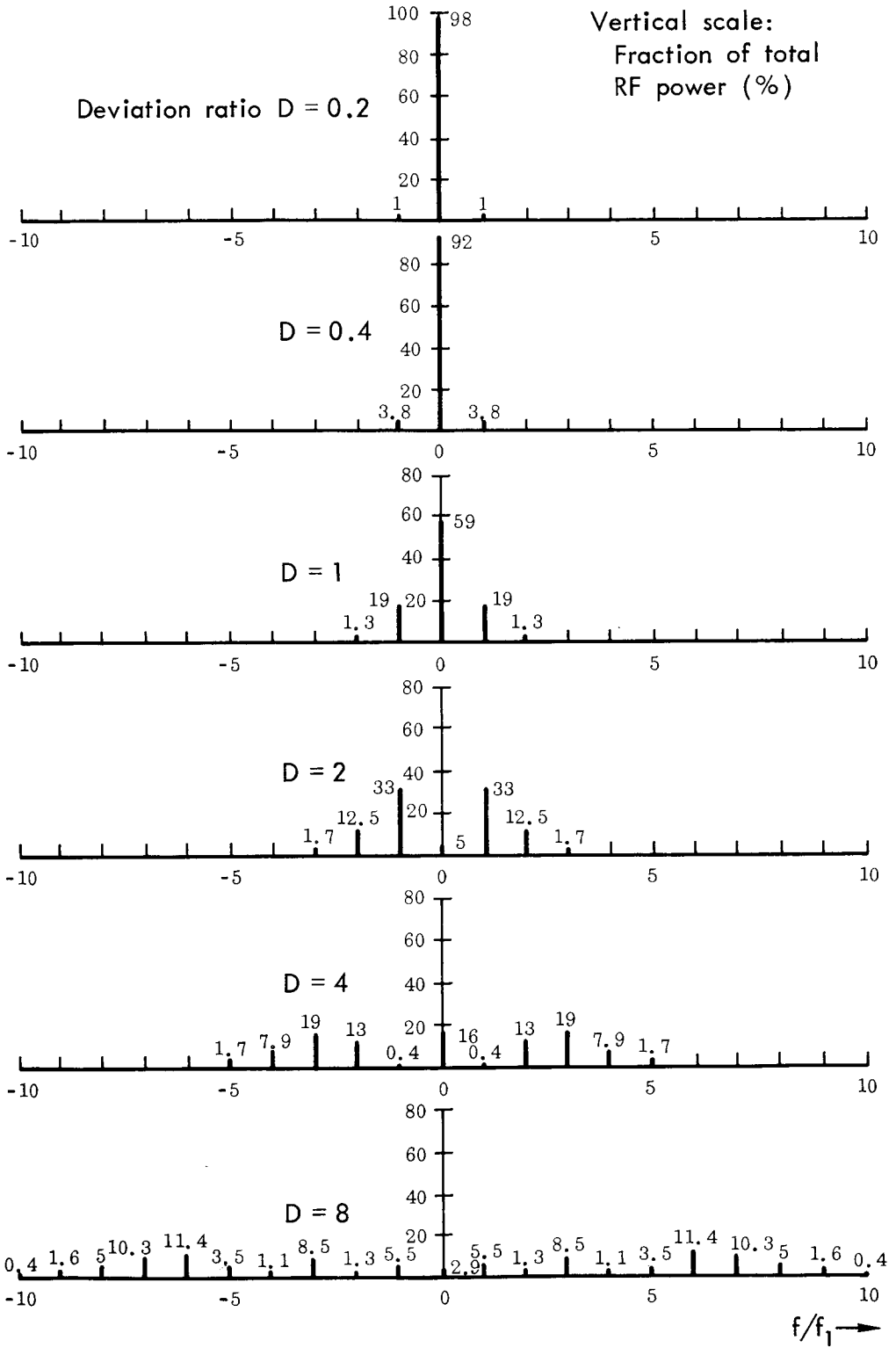


Fig. 6a—Normalized power spectra for FM by a sinusoidal test tone at frequency f_1

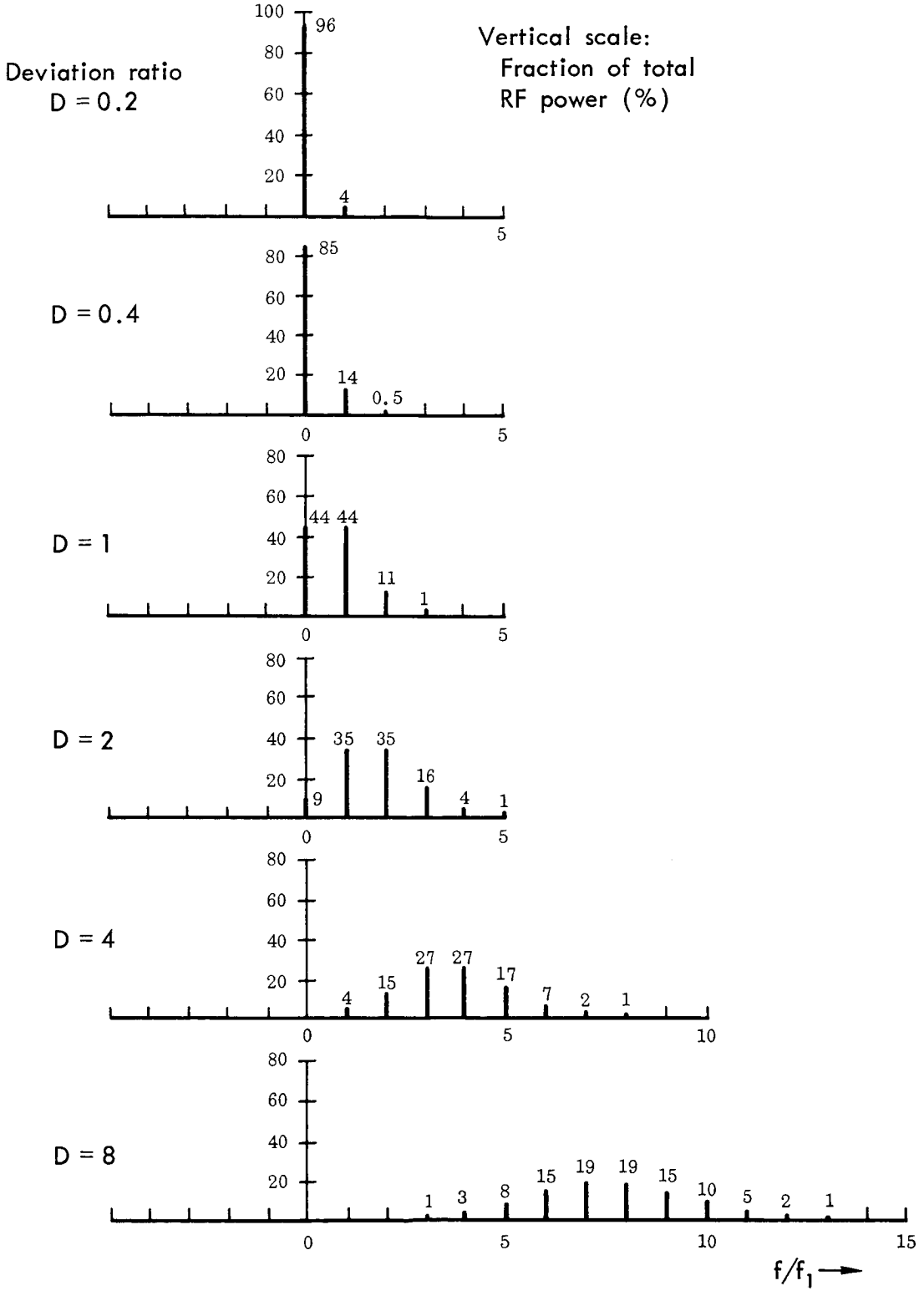


Fig. 6b—Normalized power spectra for SSBFM modulation by a sinusoidal test tone at frequency f_1

for which

$$\sum_{n=-B_R/2f_1}^{B_R/2f_1} J_n^2(D) \geq 1 - \beta \qquad \text{FM by a test tone} \qquad (115)$$

Plots of B_R/f_1 versus D for various values of β are shown in Fig. 7a on page 86.* From these curves it is apparent that B_R always exceeds $2f_1$, although for low deviation ratios (narrow-band FM) it closely approaches this value.

In the case of SSBFM, b(t) is again given by Eq. (109), but m(t) is now specified by Eq. (102). Thus, when the baseband is the test tone $e_B(t) = \cos \omega_1 t$, the instantaneous phase deviation and its Hilbert transform are respectively

$$b(t) = D \sin \omega_1 t \ , \qquad \check{b}(t) = -D \cos \omega_1 t$$

and the modulation function is

$$m(t) = e^{D(\pm \cos \omega_1 t + j \sin \omega_1 t)} = \exp\left(\pm D e^{\pm j\omega_1 t}\right) \qquad (116)$$

where the upper sign corresponds to an USB, and the lower to a LSB. Again, m(t) is periodic and deterministic so that its power spectrum is given by Eq. (111). However, in this case, the Fourier coefficients M_n may be determined by inspection upon comparing Eq. (113) with the series expansion of the exponential in Eq. (116):

$$\exp\left(\pm D e^{\pm j\omega_1 t}\right) = \sum_{n=0}^{\infty} \frac{(\pm D)^n}{n!} e^{\pm jn\omega_1 t}$$

*The points in Fig. 7a give the values of B_R/f_1 for integer values of D and the smooth curves represent a rough fit to these points. The points do not usually lie on the curves because of the discrete nature of the spectrum.

Thus, for USBFM, the Fourier coefficients are

$$M_n = \begin{cases} 0 & n < 0 \\ D^n/n! & n \geq 0 \end{cases} \tag{117}$$

while for LSBFM

$$M_n = \begin{cases} (-D)^{-n}/(-n)! & n \leq 0 \\ 0 & n > 0 \end{cases} \tag{118}$$

The one-sided character of these Fourier spectra is evident. It will also be noted that corresponding harmonics have the same amplitude; indeed, the only difference between the USB and LSB spectra is in the phase of the odd harmonics. No phase information appears in power spectra however, so the USB spectrum is simply a mirror image of the LSB spectrum. This is also apparent from the formal expressions for the power spectra obtained by combining Eq. (111) with Eqs. (117) and (118)

$$W_m(f) = \begin{cases} \sum_{n=0}^{\infty} (D^n/n!)^2 \, \delta(f - nf_1) & \text{USBFM} \\ \sum_{n=0}^{\infty} (D^n/n!)^2 \, \delta(f + nf_1) & \text{LSBFM} \end{cases} \tag{119}$$

Plots of the USB spectrum for several values of D are shown in Fig. 6b, page 83, where, to facilitate comparison with the conventional FM spectra in Fig. 6a, the SSBFM spectra have been normalized to the same total spectral power.

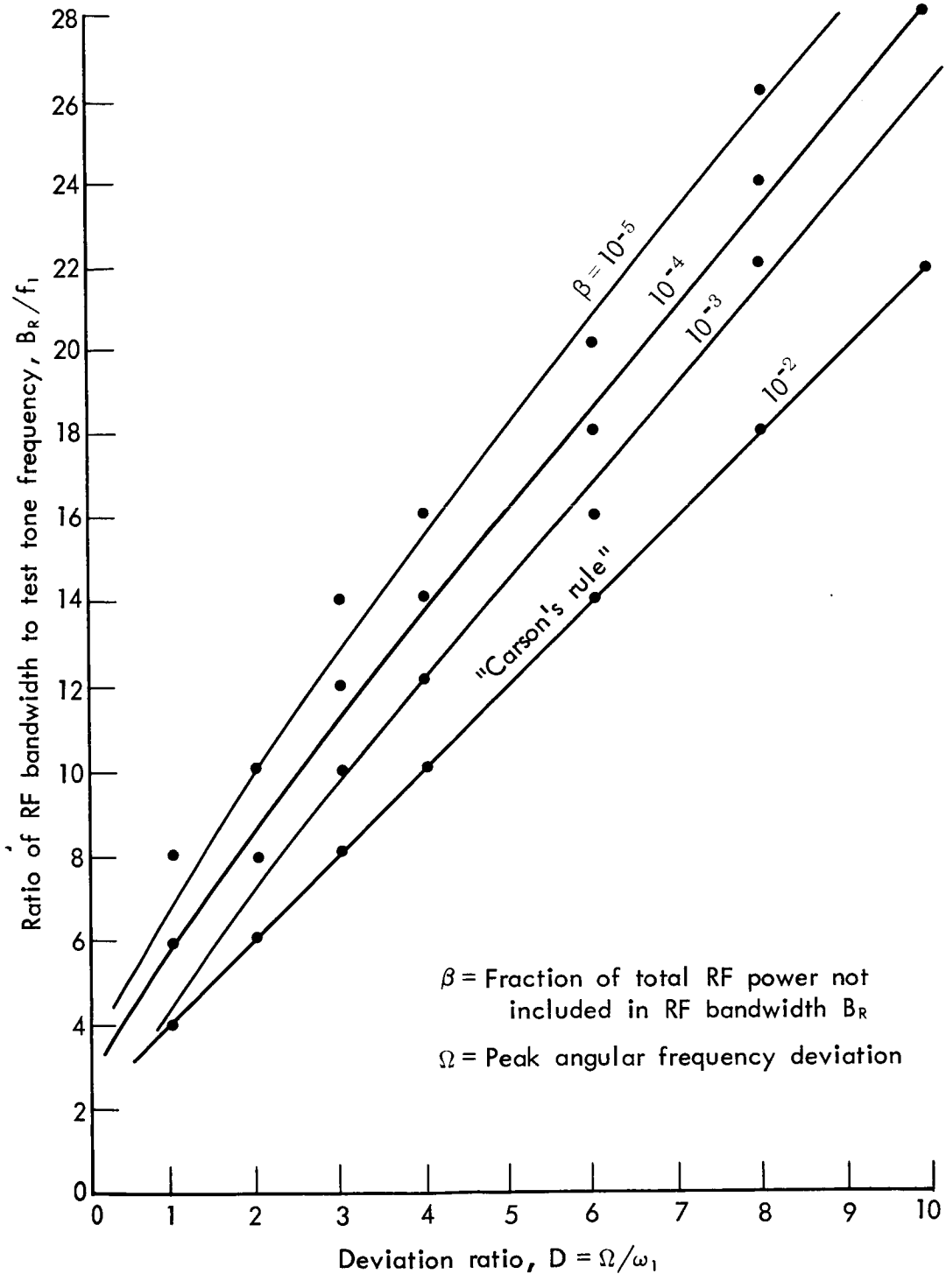


Fig. 7a—RF bandwidth for FM by a sinusoidal test tone at frequency $f_1 = \omega_1/2\pi$

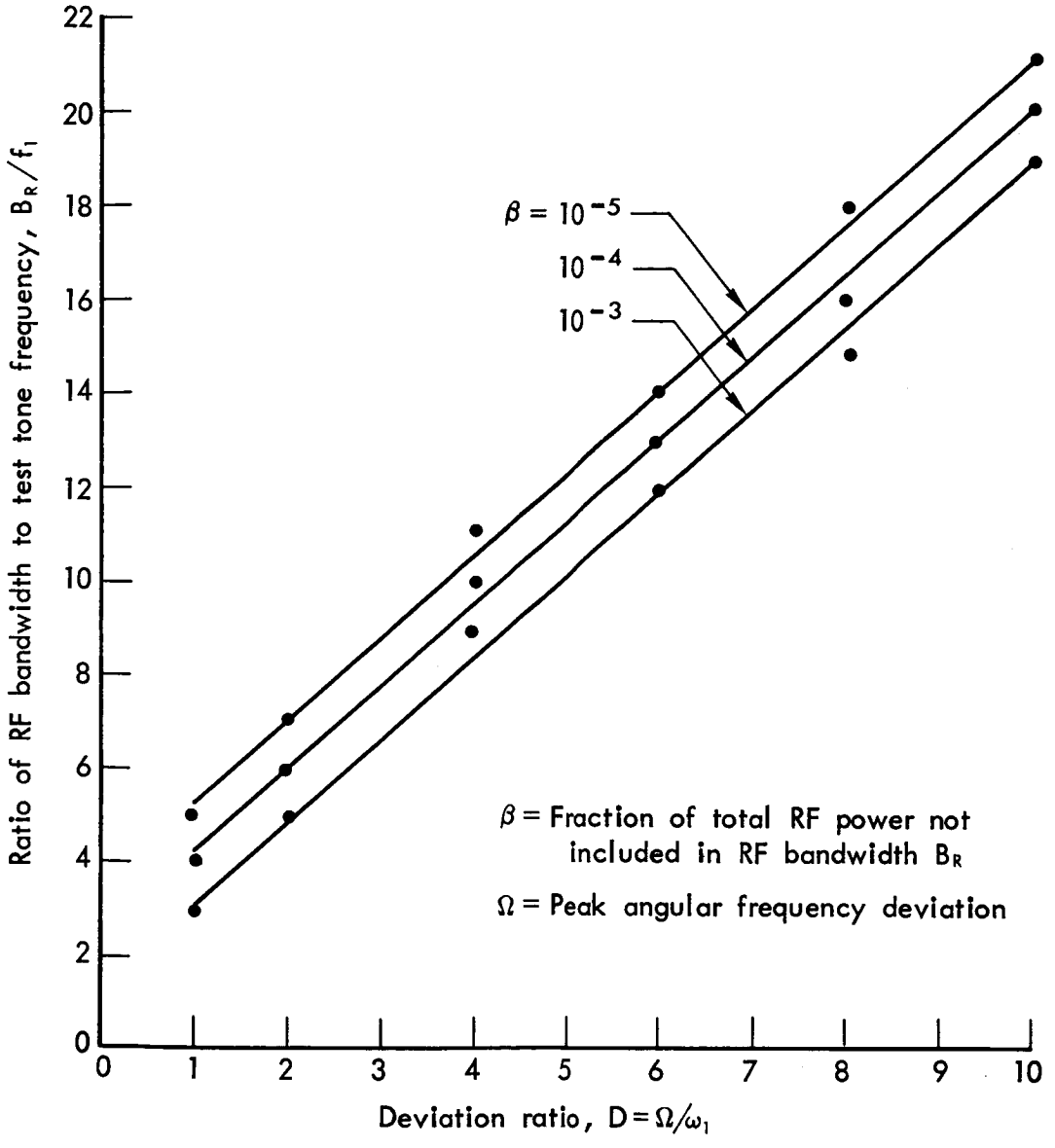


Fig. 7b—RF bandwidth for SSBFM modulation by a sinusoidal test tone frequency $f_1 = \omega_1/2\pi$

Note that there is a dc component which corresponds to a line at the carrier frequency in the power spectrum of the modulated carrier. But unlike conventional FM, where the amount of power in the carrier component decreases with increasing deviation ratio in such a way as to keep the total spectral power constant, the power of the carrier component in SSBFM is itself constant and equal to that of the unmodulated carrier. In this respect, SSBFM is similar to conventional AM, and the carrier in SSBFM could be suppressed without loss of information. When the deviation ratio exceeds unity, however, the sideband power greatly exceeds that in the carrier component and there is little motivation for carrier suppression. Moreover, carrier reinsertion might be very difficult in practice since correct amplitude, as well as frequency and phase are important.

The bandwidth for SSBFM with a test tone may be found by substituting Eq. (119) into Eq. (94) with the appropriate limits of integration and noting that the total spectral power is

$$\int_{-\infty}^{\infty} W_m(f)\, df = \sum_{n=0}^{\infty} (D^n/n!)^2 = I_o(2D)$$

in agreement with the result obtained previously using the probability density function of Eq. (80). The RF bandwidth is the smallest value of B_R for which

$$\sum_{n=0}^{B_R/f_1} (D^n/n!)^2 \geq (1 - \beta)\, I_o(2D) \qquad \text{SSBFM by test tone} \qquad (120)$$

and values of B_R/f_1 calculated from this expression are plotted* versus D in Fig. 7b, page 87. Comparing these curves with those in Fig. 7a for conventional FM, it is seen that the RF bandwidth with SSBFM is approximately one-third smaller at any given deviation ratio.

It remains to determine the power spectra and RF bandwidths for carriers that have been exponentially modulated by a gaussian distributed baseband. Using the procedure outlined in the discussion of arbitrary basebands, the first step is to obtain the characteristic function for the baseband. From this, the autocorrelation function of the modulation function is readily obtained using Eq. (101) for PM and FM, and Eq. (105) for SSBPM and SSBFM. Finally, the power spectrum of the modulation function is derived by taking the Fourier transform of the autocorrelation function.

For this purpose it will be assumed that the baseband statistics are not only gaussian in the first order (the simple probability density function hitherto considered) but in the second (bivariate) and all higher orders as well. This is equivalent to saying that the baseband is a "gaussian random process." As such, the characteristic function corresponding to the bivariate probability density function is[6]

$$C_{b_1 b_2}(\xi_1,\xi_2) \equiv e^{\overline{j(\xi_1 b_1 + \xi_2 b_2)}} = e^{-\frac{1}{2}\left(\xi_1^2\overline{b_1^2} + 2\xi_1\xi_2\,\overline{b_1 b_2} + \xi_2^2\overline{b_2^2}\right)} \tag{121a}$$

and the fourth-order characteristic function is[6]

$$C_{b_1 b_2 b_3 b_4}(\xi_1,\xi_2,\xi_3,\xi_4) \equiv e^{j\overline{\sum_{i=1}^{4}\xi_i b_i}} = e^{-\frac{1}{2}\sum_{r=1}^{4}\sum_{s=1}^{4}(\xi_r\xi_s\overline{b_r b_s})} \tag{121b}$$

*See footnote on page 84.

where

$$b_1 = b(t),\ b_2 = b(t+\tau),\ b_3 = \breve{b}(t),\ b_4 = \breve{b}(t+\tau) \tag{121c}$$

In the case of PM and FM, Eqs. (101) and (121a) combine to give

$$R_m(\tau) = C_{b_1 b_2}(-1,\ 1) = e^{-\frac{1}{2}\left(\overline{b_1^2} - \overline{2b_1 b_2} + \overline{b_2^2}\right)}$$

But from Eq. (121c), the averages in the exponent may all be expressed in terms of $R_b(\tau)$, the autocorrelation of the baseband function:

$$\overline{b_1 b_2} = \overline{b(t)\ b(t+\tau)} \equiv R_b(\tau)$$

$$\overline{b_1^2} = \overline{b_2^2} = \overline{b^2(t)} = R_b(0)$$

With these substitutions, the relation between $R_m(\tau)$ and $R_b(\tau)$ is obtained

$$R_m(\tau) = e^{-R_b(0) + R_b(\tau)} \qquad \text{PM, FM by gaussian baseband} \tag{122a}$$

Taking the Fourier transforms of both sides yields a formal expression for the power spectrum

$$W_m(f) = e^{-R_b(0)} \int_{-\infty}^{\infty} e^{R_b(\tau) - j2\pi f\tau}\, d\tau \qquad \text{PM, FM by gaussian baseband} \tag{122b}$$

The spectra for PM and FM are distinguished by the dependence of the baseband function b(t) on the baseband. Thus from Table 2

$$b(t) = \begin{cases} \Phi e_B(t) & \text{PM} \\ \Omega \int e_B(t)\, dt & \text{FM} \end{cases} \tag{122c}$$

where Φ and Ω are respectively the peak deviations of phase and instantaneous frequency. Since b(t) is real in both cases, it follows from Eq. (60b) that $\exp\left[R_b(\tau)\right]$ is an even function of τ, and so Eq. (122b) can also be written

$$W_m(f) = 2e^{-R_b(0)} \int_0^\infty e^{R_b(\tau)} \cos 2\pi f\tau \, d\tau \qquad \text{PM, FM by gaussian baseband} \qquad (122d)$$

In the case of SSBPM and SSBFM, Eqs. (121b) and (121c) may be combined with Eq. (105) to yield

$$\begin{aligned} R_m(\tau) = \exp\Big\{-\tfrac{1}{2}\Big[& R_b(0) - R_b(\tau) \mp jR_{b\check{b}}(0) \mp jR_{b\check{b}}(\tau) \\ & - R_b(\tau) + R_b(0) \pm j R_{\check{b}b}(\tau) \pm jR_{\check{b}b}(0) \\ & \mp jR_{b\check{b}}(\tau) \pm jR_{\check{b}b}(\tau) - R_{\check{b}}(0) - R_{\check{b}}(\tau) \\ & \mp jR_{bb}(\tau) \pm jR_{\check{b}b}(0) - R_{\check{b}}(\tau) - R_{\check{b}}(0)\Big]\Big\} \end{aligned}$$

where the double subscripts indicate the cross-correlations defined in Eq. (89a). With the aid of the general relations of Eqs. (89b) - (89d), this simplifies to

$$R_m(\tau) = e^{2\left[R_b(\tau) \pm j\check{R}_b(\tau)\right]} \qquad \text{SSBPM, SSBFM by gaussian baseband} \qquad (123a)$$

Note that with the upper sign (USB), the bracketed term in the exponent is just the analytic signal representation $\underline{R}_b(\tau)$ of $R_b(\tau)$ while with the lower sign (LSB), it is $\underline{R}_b^*(\tau)$. Taking Fourier transforms gives

$$W_m(f) = \int_{-\infty}^{\infty} e^{2[R_b(\tau) \pm j\check{R}_b(\tau)] - j2\pi f\tau} d\tau \qquad \text{SSBPM, SSBFM by gaussian baseband} \quad (123b)$$

where, as in Eq. (122c), the distinction between SSBPM and SSBFM is determined by the dependence of b(t) on the baseband:

$$b(t) = \begin{cases} \Phi e_B(t) & \text{SSBPM} \\ \Omega \int e_B(t)dt & \text{SSBFM} \end{cases} \quad (123c)$$

Eqs. (122b) and (123b) show how the power spectrum $W_m(f)$ depends on $R_b(\tau)$; it remains to show how $R_b(\tau)$ depends on $W_{e_b}(f)$, the power spectrum of the baseband. Using Eq. (58) to express $R_b(\tau)$ in terms of $W_b(f)$ and then applying Eq. (62a) to Eqs. (122c) and (123c), the desired relations are

$$R_b(\tau) = \begin{cases} \Phi^2 \int_{-\infty}^{\infty} W_{e_B}(f)\, e^{j2\pi f\tau} df & \text{PM, SSBPM} \quad (124a) \\ \left(\frac{\Omega}{2\pi}\right)^2 \int_{-\infty}^{\infty} f^{-2}\, W_{e_B}(f)\, e^{j2\pi f\tau}\, df & \text{FM, SSBFM} \quad (124b) \end{cases}$$

Taking Eqs. (124) in conjunction with Eqs. (122b) and (123b), the power spectrum $W_m(f)$ corresponding to PM, FM, SSBPM, or SSBFM by any gaussian distributed baseband can be determined from a knowledge of the baseband power spectrum and the modulation index. For most basebands, the required integrations cannot be carried out in closed form and it is necessary to use approximations.[10,12,14]

One class of gaussian distributed basebands for which the exact calculation can be carried out is that for which the power spectra have the form

$$W_{e_B}(f) = cf^2 \exp\left(-d|f|\right)$$

where c and d are positive constants. Imposing the general condition

$$\int_{-\infty}^{\infty} W_{e_B}(f)df = \overline{e_B^2(t)} = \Lambda_B^{-1}$$

and letting f_p denote the frequency at which $W_{e_B}(f)$ has its maximum value, this spectrum becomes

$$W_{e_B}(f) = \frac{2}{f_p \Lambda_B}\left(\frac{f}{f_p}\right)^2 \exp\left(-2\frac{|f|}{f_p}\right)$$

Substituting it into Eq. (124b) yields the autocorrelation of the base-band function (or phase deviation) for FM and SSBFM

$$R_b(\tau) = 8\left(\frac{f_d}{f_p}\right)^2 \frac{1}{4 + (\omega_p\tau)^2}$$

where $\omega_p = 2\pi f_p$, and

$$f_d^2 = (\Omega/2\pi)^2/\Lambda_B = B_B^2 D^2/\Lambda_B$$

is the mean square frequency deviation. The Hilbert transform of $R_b(\tau)$ is easily found using the definition of Eq. (7)

$$\breve{R}_b(\tau) = 4\left(\frac{f_d}{f_p}\right)^2 \frac{\omega_p\tau}{4 + (\omega_p\tau)^2}$$

Finally, the power spectra of the modulation functions for FM and SSBFM modulated carriers can be calculated by substituting the expressions for $R_b(\tau)$ and $\breve{R}_b(\tau)$ into Eqs. (122d) and (123b) respectively.

The resultant values* of $W_m(f)$ are plotted versus f/f_p for various deviation ratios D in Fig. 8a in the case of FM, and Fig. 8b in the

*Personal communication with E. Bedrosian of The RAND Corporation.

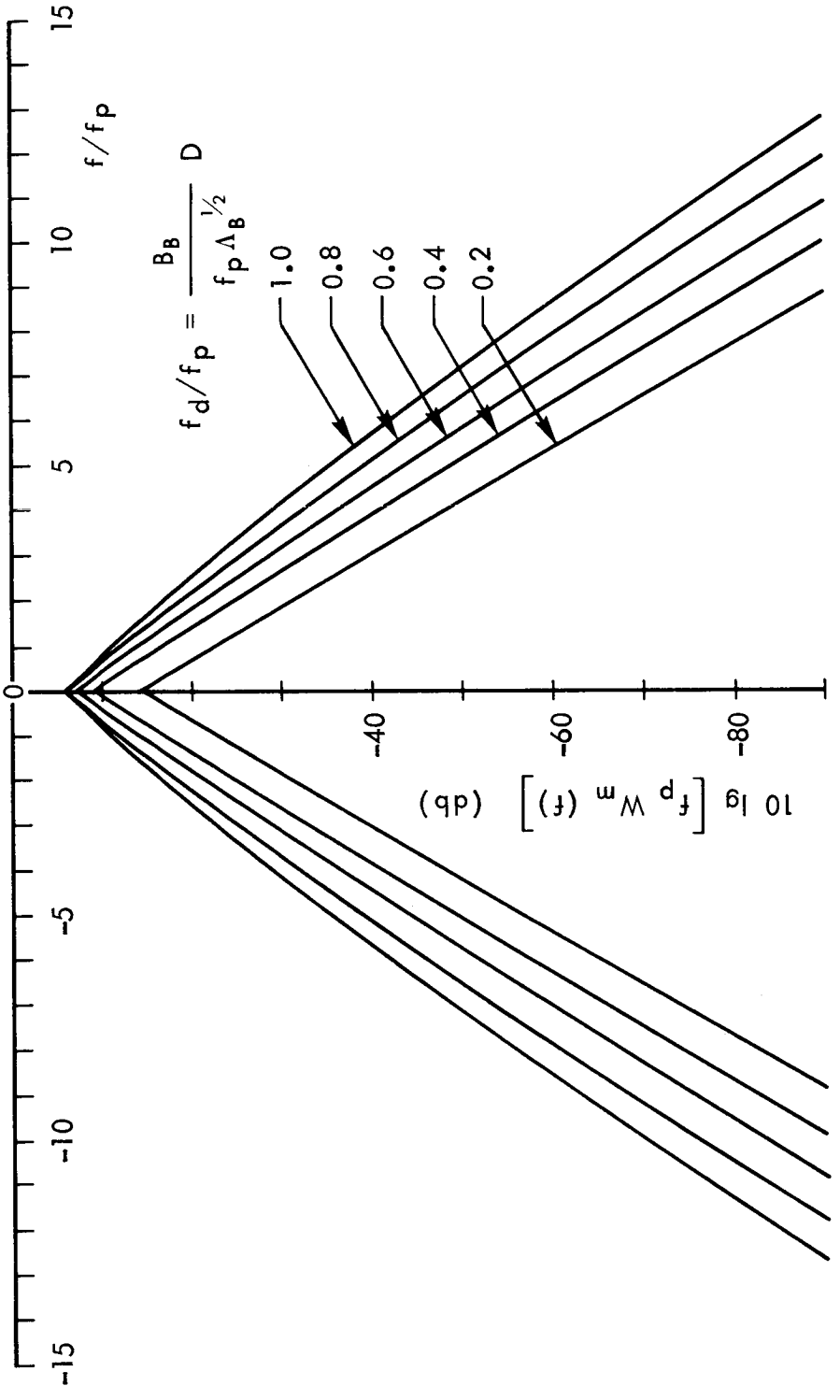


Fig. 8a — Power spectra for FM by a gaussian baseband with power spectrum

$$W_{e_B}(f) = \frac{2}{f_p \Delta_B} \left(\frac{f}{f_p}\right)^2 e^{-2|f|/f_p}$$

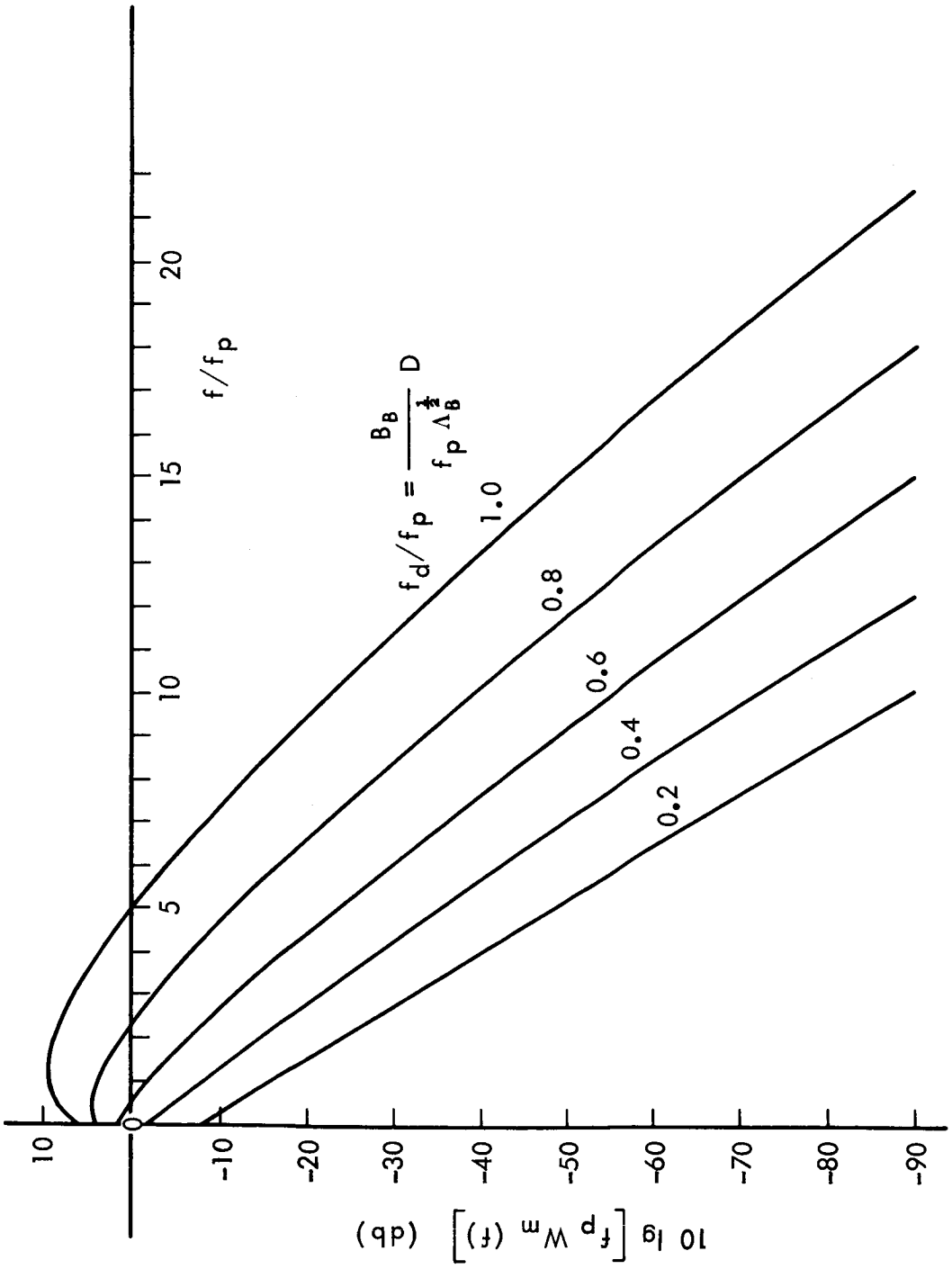


Fig. 8b—Power spectra for SSBFM by a gaussian baseband with power spectrum

$$W_{e_B}(f) = \frac{2}{f_p \Delta_B} \left(\frac{f}{f_p}\right)^2 e^{-2|f|/f_p}$$

case of SSBFM. Note that, as in the case of a sinusoidal baseband, the RF bandwidth for SSBFM is roughly two-thirds that for FM.

With regard to RF bandwidth, however, it should be noted that for each of the exponential modulation methods, the value of B_R defined in Eq. (94) depends not only on the basebandwidth and the modulation index but on the shape of the baseband power spectrum as well. Thus, even with specified values of β and B_B, no general result can be given for the ratio B_R/B_B. Nonetheless, for FM a useful approximation applicable to a variety of basebands and a wide range of modulation indices, D, is given by Carson's Rule[10,11]*

$$B_R = 2B_B(1 + D) \qquad \text{FM} \qquad (125a)$$

For SSBFM it will be assumed that the RF bandwidth is two-thirds as large--i.e.,

$$B_R \simeq \frac{4}{3} B_B(1 + D) \qquad \text{SSBFM} \qquad (125b)$$

Under certain circumstances, Carson's Rule can also be applied to obtain an expression for the bandwidth of a phase modulated carrier in terms of the peak phase deviation or PM modulation index, Φ. Thus, if in addition to the power spectrum and probability density function of the baseband, the probability density of the derivative of the baseband signal is known, the peak instantaneous frequency deviation, and hence the corresponding deviation ratio, can be calculated as a function of Φ. For example, if the baseband has a gaussian amplitude distribution and a flat low-pass power spectrum, the amplitude distribution of the derivative of the baseband will also be gaussian,

*Note from Fig. 7a that for a sinusoidal modulating signal, Carson's rule corresponds to $\beta = 10^{-2}$.

and it is easily shown that the deviation ratio corresponding to PM with index Φ is simply $\Phi/\sqrt{3}$. Hence the RF bandwidth of such a PM wave is

$$B_R = 2B_B(1 + \Phi/\sqrt{3}) \qquad \text{PM by low-pass flat gaussian baseband} \qquad (125c)$$

The foregoing expressions for B_R have been included in the last column of Table 3 with the qualification that for greater accuracy, the RF bandwidth should be calculated from the RF power spectrum $W_{E_R}(f)$ whenever it is available.

NOISE AND DEMODULATION

Introduction

So far in this section, attention has been devoted exclusively to the modulation process, and in particular, to a statistical description of the waveform that results when a sinusoidal carrier is analog modulated by an information carrying baseband. The next step is to consider the inverse process of recovering the baseband from the modulated carrier. This process takes place in the receiver, which is assumed to be of the superheterodyne type consisting of an IF (intermediate frequency) section followed by a demodulator.

Except in the case of receivers employing feedback loops, the IF section is not directly involved in signal recovery; its operation is essentially the same regardless of the type of modulation for which the receiver is designed. The principal components are a mixer or converter followed by an IF amplifier. The converter serves to shift the frequency of the received signal by multiplying it with a locally

generated sinusoid whose angular frequency differs from that of the unmodulated carrier by a fixed amount ω_I called the intermediate frequency. For example, with the general received signal of Eq. (5)

$$E_R(t) = A_R\, e(t) = a(t)\, A_R \cos\left[\omega_o t + \varphi(t)\right]$$

multiplication by the local sinusoid $\cos(\omega_o - \omega_I)t$ yields

$$E_R(t)\cos(\omega_o - \omega_I)t = \frac{1}{2}\, a(t)\, A_R \cos\left[\omega_I t + \varphi(t)\right] + \frac{1}{2}\, a(t)\, A_R \cos\left[(2\omega_o - \omega_I)t + \varphi(t)\right]$$

The converter output thus consists of two terms, each similar to the received signal except for the replacement of ω_o by ω_I and $2\omega_o - \omega_I$ respectively. With ω_I chosen so that the spectra of these terms do not overlap, the IF amplifier then band-pass filters and amplifies the converter output so as to retain only the first term. Assuming the IF filters to be ideally rectangular with bandwidth equal to the RF bandwidth B_R of the modulated carrier, the signal at the output of the IF section is just an undistorted replica of the first term:

$$E_I(t) = a(t)\, A_I \cos\left[\omega_I t + \varphi(t)\right] \tag{126}$$

Here A_I represents the amplitude of the IF amplifier output when the unmodulated carrier $[a(t) = 1,\ \varphi(t) = 0]$ is transmitted, A_I/A_R is the overall signal gain of the IF section, and the phase constant representing the transmission delay of this section is neglected. Thus, when normalized to A_I, the signal input to the demodulator,

$$e_I(t) \equiv E_I(t)/A_I = a(t) \cos \left[\omega_I t + \varphi(t)\right]$$

differs only in frequency from the normalized modulated carrier e(t) defined in Eq. (6) by the amplitude ratio a(t) and phase deviation $\varphi(t)$ imposed at the transmitter.

Unfortunately, this is not the complete waveform that arrives at the input to the demodulator section of the receiver. The signal $e_I(t)$ is inevitably accompanied by additive noise* --i.e., a random waveform representing the combined effect of all the unwanted sources of RF energy having an electrical path to the demodulator. These noise sources include terrestrial and extraterrestrial emitters within the receiving antenna beam, the transmission lines from the antenna to the receiver, and especially the input stages (front end) of the receiver IF section. It is from the combination of modulated signal and noise that the demodulator section of the receiver must recover the baseband signal. So, before discussing the operation of various types of demodulators, the mathematical representation of noise waveforms will be considered.

Description of Noise

In describing the additive noise with which the modulated carrier must compete, it will be assumed that there is no interference in the form of other modulated carriers or impulsive noise from man-made sources in the receiver pass band. Under these circumstances, the additive noise

*The amplitude and phase of the carrier may also be affected by nonlinearities in the transmitting and receiving amplifiers and by time-dependent absorption, multiple scattering, and other processes in the propagation medium. These sources of distortion and so-called multiplicative noise are here assumed to be negligible. In cases where they are not, the methods developed in Ref. 15 may be applied.

waveform at the demodulator input has a gaussian amplitude distribution and a power spectrum that is essentially flat (independent of frequency) except as modified by the filters of the IF amplifier. Since these filters were assumed to have a bandwidth B_R that is nearly always small compared with the IF frequency, the noise input to the demodulator is narrow band.

Any narrow-band waveform may be viewed as the result of simultaneously modulating the envelope and phase of a sinusoidal carrier whose frequency lies in the narrow band.[11] In particular, if the noise waveform at the demodulator input is $U_I(t)$, there exists an envelope ratio $v(t)$ and phase deviation $\psi(t)$ such that

$$U_I(t)/A_I = v(t) \cos \left[\omega_I t + \psi(t)\right] \equiv u_I(t) \tag{127}$$

where ω_I is the receiver IF frequency, and A_I is the unmodulated carrier amplitude at the demodulator input.

Alternatively, the normalized input noise waveform $u_I(t)$ may be viewed as the difference between DSB modulated quadrature carriers:

$$u_I(t) = p(t) \cos \omega_I t - q(t) \sin \omega_I t \tag{128}$$

where

$$\begin{aligned} p(t) &= v(t) \cos \psi(t) \\ q(t) &= v(t) \sin \psi(t) \end{aligned} \tag{129}$$

are respectively called the in-phase and quadrature noise components.

In either case, it is customary to regard the noise at the demodulator input as the result of a noise waveform $U_R(t)$ at the input of an idealized noise-free receiver. By analogy with the relation between $E_R(t)$ and $e_I(t)$, it follows that

$$U_R(t)/A_R \equiv u_R(t) = v(t) \cos \left[\omega_o t + \psi(t)\right]$$
$$= p(t) \cos \omega_o t - q(t) \sin \omega_o t \tag{130}$$

where A_R is the unmodulated carrier amplitude at the receiver input, and any initial phase constant has been neglected.

The statistical characteristics of $U_R(t)$ are of course the same as those already mentioned for the noise at the demodulator input--i.e., a gaussian probability density

$$p_{U_R}(x) = \frac{1}{\sqrt{2\pi}\,\sigma_{U_R}} \exp\left(-x^2/2\sigma_{U_R}{}^2\right) \tag{131}$$

and a rectangular power spectrum

$$W_{U_R}(f) = \begin{cases} N_o/2 & f \text{ in RF bandwidth} \\ 0 & \text{otherwise} \end{cases} \tag{132}$$

where $\sigma_{U_R} = \sqrt{N_R}$ is the rms value of the noise waveform at the receiver input, and N_o is the constant noise power density within the RF band pass at this point. These two quantities are related through the various equivalent expressions for the input noise power:

$$N_R \equiv \overline{U_R^2(t)} = \sigma_{U_R}^2 = \int_{-\infty}^{\infty} W_{U_R}(f)df = N_o B_R \tag{133}$$

The reason for referring the noise at the demodulator input back to the receiver input is that the latter point is the one generally used for specifying the total noise power, even though in practice, contributions to the output noise often arise both before and after the receiver input. The specification is usually made in terms of a

system noise temperature T_s such that the input noise power density is

$$N_o = k\ T_s \tag{134}$$

where k is Boltzmann's constant ($k = 1.38 \times 10^{-23}\ \frac{\text{joules}}{\text{degree}}$). Thus, another expression for the total noise power referred to the receiver input is

$$N_R = k\ T_s\ B_R$$

The statistical characteristics of the modulations v(t) and $\psi(t)$ which characterize the noise waveforms may be derived from those just given for $U_R(t)$. The envelope ratio v(t) is Rayleigh distributed[11] with probability density

$$p_v(x) = \frac{x}{\sigma^2_{u_R}} \exp\ (-x^2/2\sigma^2_{u_R}) \qquad x \geq 0 \tag{135}$$

where σ_{u_R} is the rms value of the normalized input noise, and the phase deviation $\psi(t)$ is uniformly distributed

$$p_\psi(x) = \frac{1}{2\pi} \qquad 0 \leq x \leq 2\pi \tag{136}$$

From the distribution for v(t) it follows at once that

$$\overline{v^2(t)} = 2\sigma^2_{u_R} \equiv \overline{2u^2_R(t)} = \overline{2U^2_R(t)}/A^2_R = N_R/S_{Ro} \tag{137}$$

where

$$S_{Ro} = A^2_R/2$$

is the power of the unmodulated carrier at the receiver input. Thus, the mean square envelope ratio is just the total input noise power N_R

normalized to the average power S_{Ro} of the unmodulated carrier. The distributions for $v(t)$ and $\psi(t)$ may also be applied to calculate the mean square values of the noise components $p(t)$ and $q(t)$ defined in Eq. (129) with the result[11]

$$\overline{p^2(t)} = \overline{q^2(t)} = \frac{1}{2}\,\overline{v^2(t)} = \frac{1}{2}\,N_R/S_{Ro} \tag{138}$$

Having described the input noise, the problem is to determine for each type of modulation, the fidelity with which the baseband signal can be recovered by an appropriate demodulator from the mixture of signal and noise at the demodulator input. Fidelity of recovery will be measured by the ratio K_L of average signal power S_L to average noise power N_L at the output of the low-pass filter which forms the output stage of most demodulators. The proportions of signal and noise in the additive signal-noise mixture at the demodulator input will be measured by the ratio K_R of average signal power S_R to average noise power N_R at the receiver input. In this connection, it will be assumed in all cases that the signal power greatly exceeds the noise power

$$K_R \gg 1$$

To determine the relation between K_L and K_R, the composite signal plus noise waveform $X_I(t)$ at the demodulator input will be written in normalized form

$$X_I(t)/A_I = x_I(t) = e_I(t) + u_I(t) = r(t)\cos\left[\omega_I t + \theta(t)\right] \tag{139}$$

where $e_I(t)$ and $u_I(t)$ are given by Eqs. (126) and (127), and referring to the vector diagram of Fig. 9, the envelope ratio $r(t)$ and phase deviation $\theta(t)$ are given in terms of the corresponding signal and noise quantities by

$$r(t) = \sqrt{a^2(t) + 2\,a(t)\,v(t)\cos\left[\psi(t) - \varphi(t)\right] + v^2(t)} \tag{140}$$

$$\theta(t) = \cos^{-1}\frac{a(t)\cos\varphi(t) + v(t)\cos\psi(t)}{r(t)} \tag{141}$$

It will be noted that in the absence of noise, $r(t)$ and $\theta(t)$ respectively reduce to $|a(t)|$ and $\varphi(t)$, as they should. Also note that, with noise present, the waveform $x_I(t)$ at the demodulator input is still modulated in both amplitude and phase--even when the input signal $e_I(t)$ is modulated in only one (or neither). Moreover, as will next be shown for each of the important types of analog demodulators in turn, the waveform at the demodulator output can, with one important qualification, be expressed quite simply in terms of $r(t)$ and/or $\theta(t)$.

Product Demodulator

An ideal product demodulator (also called a synchronous, coherent, or homodyne detector) is one in which the input waveform $x_I(t)$ is first multiplied by a locally generated replica of the unmodulated sinusoidal carrier. The amplitude of this local carrier is not important, but, for distortion-free demodulation, the frequency and phase of the local carrier should be the same as those of the input signal $e_I(t)$ in the absence of modulation; the local carrier is thus taken to be simply $\cos \omega_I t$. The

Unmodulated carrier waveform	$\text{Re}\ e^{j\omega_I t} = \cos \omega_I t$
Signal waveform	$e_I(t) = \text{Re}\ \underline{e}_I(t) = a(t) \cos [\omega_I t + \varphi(t)]$
Noise waveform	$u_I(t) = \text{Re}\ \underline{u}_I(t) = v(t) \cos [\omega_I t + \psi(t)]$
Signal plus noise waveform	$x_I(t) = \text{Re}\ \underline{x}_I(t) = r(t) \cos [\omega_I t + \theta(t)]$

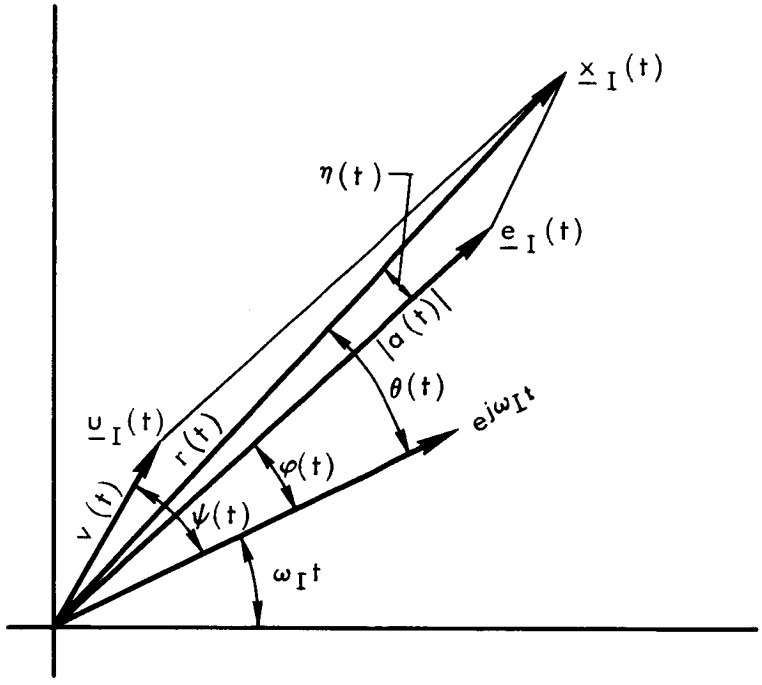


Fig. 9—Rotating vector representation of normalized waveforms at the demodulator input when input SNR $K_R >> 1$

output $x_L(t)$ of the product demodulator is then obtained by passing the product, $x_I(t) \cos \omega_I t$, through a low-pass filter whose bandwidth is at least equal to the RF bandwidth B_R but less than the IF frequency ω_I.*

A quite general expression for $x_L(t)$ can be obtained as follows. Let $\underline{x}_I(t)$ be the analytic signal representation of $x_I(t)$

$$\underline{x}_I(t) = x_I(t) + j\check{x}_I(t) \tag{142}$$

Then the signal-plus-noise input waveform may be written

$$x_I(t) = \frac{1}{2}\left[\underline{x}_I(t) + \underline{x}_I^*(t)\right]$$

where the spectrum of the first term contains only positive frequencies and is identical to the positive frequency spectrum of $2x_I(t)$, and the spectrum of the second term contains only negative frequencies and is equal to the negative frequency spectrum of $2x_I(t)$. Since $x_I(t)$ is assumed to be narrow band with bandwidth B_R, this means that the spectrum of $\underline{x}_I(t)$ is confined to the interval $f_I \pm B_R$ and that of $\underline{x}_I^*(t)$ to the interval $-f_I \pm B_R$.** Writing the local carrier

$$\cos \omega_I t = \frac{1}{2} e^{j\omega_I t} + \frac{1}{2} e^{-j\omega_I t}$$

---

*In practice, this bandwidth need be no greater than $B_R/2$ for double-sidebanded signals and B_R for single-sidebanded signals.

**These are generous bounds. For double-sidebanded signals, $\underline{x}_I(t)$ actually occupies the frequency range $f_I \pm B_R/2$. For single-sidebanded signals, the range is $f_I \leq f \leq f_I + B_R$ if the sideband lies above the carrier, and $f_I - B_R \leq f \leq f_I$ if below.

the signal product becomes

$$x_I(t) \cos \omega_I t = \frac{1}{4}\left[e^{-j\omega_i t}\, \underline{x}_I^*(t) + e^{-j\omega_I t}\, \underline{x}_I(t) + e^{j\omega_I t}\, x_I^*(t) + e^{j\omega_I t}\, \underline{x}_I(t)\right] \tag{143}$$

To determine the effect of the ideal low-pass filter on this product, recall that multiplying a time-function by $e^{j\omega_I t}$ merely shifts its spectrum to higher frequencies by an amount ω_I, while multiplication by $e^{-j\omega_I t}$ shifts its spectrum downward by the same amount. Thus the spectra of the first and fourth terms of Eq. (143) are respectively confined to the frequency ranges $-2f_I \pm B_R$ and $+2f_I \pm B_R$, while the spectra of the second and third terms lie in the range $\pm B_R$. Thus the output $x_L(t)$ of the low-pass filter consists of just the second and third terms

$$x_L(t) = \frac{1}{4}\left[e^{-j\omega_I t}\, \underline{x}_I(t) + e^{j\omega_I t}\, \underline{x}_I^*(t)\right]$$

or, substituting from Eq. (142),

$$x_L(t) = \frac{1}{2}\left[x_I(t) \cos \omega_I t + \breve{x}_I(t) \sin \omega_I t\right] \tag{144}$$

Now, providing that $r(t)$ and $\theta(t)$ are the envelope and phase deviation defined in terms of the analytic signal representation of $x_I(t)$,

$$r(t) = |\underline{x}_I(t)| = \sqrt{x_I^2(t) + \breve{x}_I^2(t)} \tag{145}$$

$$\omega_I t + \theta(t) = \arg \underline{x}_I(t) = \tan^{-1}\left[\check{x}_I(t)/x_I(t)\right] \tag{146}$$

the input signal plus noise and its Hilbert transform may be written

$$x_I(t) = r(t) \cos\left[\omega_I t + \theta(t)\right] \tag{147}$$

$$\check{x}_I(t) = r(t) \sin\left[\omega_I t + \theta(t)\right] \tag{148}$$

Substituting these expressions into Eq. (144) yields the desired expression for the output of a product demodulator in terms of the envelope and phase deviation which characterize the signal plus noise waveform at its input:

$$x_L(t) = \frac{1}{2} r(t) \cos \theta(t) \tag{149}$$

It might be noted from Fig. 9 that $x_L(t)$ is just the component of the analytic signal representation of the input $\underline{x}_I(t)$ in the direction of the carrier phasor $e^{j\omega_I t}$; i.e., the component of $\underline{x}_I(t)$ that is "in phase" with the carrier.

The simplicity of this result underscores the advantage of defining $r(t)$ and $\theta(t)$ in terms of the analytic signal $\underline{x}_I(t)$ --i.e., by means of Eqs. (145) and (146). At the same time, it should be recognized that, so far as the input waveform $x_I(t)$ is concerned, it could in principle have been produced by simultaneously modulating the amplitude and phase of a carrier $\cos \omega_I t$ using any combination of amplitude ratio $r'(t)$ and phase deviation $\theta'(t)$ for which

$$r'(t) \cos\left[\omega_I t + \theta'(t)\right] \equiv x_I(t) \tag{150}$$

There are, in fact, infinitely many such combinations; for example, either $r'(t)$ or $\theta'(t)$ may be chosen arbitrarily, and Eq. (150) solved for the other.*

The significant point is that, given a particular input waveform $x_I(t)$, the output of a product demodulator does not depend on which equivalent combination of $r'(t)$ and $\theta'(t)$ was used to produce $x_I(t)$ (or to represent it mathematically). Rather, as confirmed in Eq. (149), it depends only on the characteristics of $x_I(t)$ itself. The advantage of defining the envelope $r(t)$ and phase deviation $\theta(t)$ in terms of the analytic signal is that they then represent precisely those characteristics of a narrow-band input signal in terms of which the outputs of ideal product demodulators can be expressed completely and compactly. Mathematically, this is a consequence of the fact that, of all the combinations $r'(t)$, $\theta'(t)$ which meet condition (150), $r(t)$ and $\theta(t)$ represent the only pair which also meets the condition (148) on which the simplicity of Eq. (149) depends.

Use of the analytic signal to define the envelope $r(t)$ and phase deviation $\theta(t)$ of the input waveform is thus the important qualification to the assertion that the output of the various ideal demodulators can be expressed simply in terms of $r(t)$ and/or $\theta(t)$. Indeed, this property of the magnitude and argument of the analytic signal makes the analytic signal representation as important to the theory of demodulation as it is coming to be recognized in the theory of modulation. As a consequence, the terms "envelope" and "phase deviation" of a signal

*However note that if $\theta'(t)$ causes $\cos[\omega_I t + \theta'(t)]$ to vanish at a time when $x_I(t) \neq 0$, an infinite value of $r'(t)$ will be required at that instant.

will henceforth be restricted here to the functions defined as in Eqs. (145) and (146) by the analytic signal representation of the signal.

Returning now to the calculation of the signal-to-noise power ratio at the output of a product demodulator, it is first necessary to define what is meant by the signal and the noise components of the output waveform $x_L(t)$. For this purpose, $x_L(t)$ is expressed in terms of the amplitude ratios and phase deviations of the input signal and noise by substituting Eq. (141) into Eq. (149):

$$x_L(t) = \frac{1}{2} a(t) \cos \varphi(t) + \frac{1}{2} v(t) \cos \psi(t) \tag{151}$$

The first thing to note about this expression is that it divides into two terms, one of which depends only on the input signal $e_I(t)$, and the other only on the input noise $u_I(t)$. This illustrates the fact that the product detector is linear in the sense that, with a sum of signals at the input, the output is the sum of the outputs that would be produced by each input alone. Regardless of the relative magnitudes of the signal and noise inputs, the signal-dependent output term remains separate from and hence undistorted by the noise-dependent term.

The noise-dependent term is the response of the product demodulator to the normalized noise input $u_I(t)$ and will be denoted $u_L(t)$:

$$u_L(t) = \frac{1}{2} v(t) \cos \psi(t) \equiv \frac{1}{2} p(t) \tag{152}$$

The identification in terms of $p(t)$, the in-phase component of $u_I(t)$, follows from Eq. (129). The demodulator output corresponding to the total input noise waveform $U_I(t) = A_I u_I(t)$ is

$$U_L(t) = A_I u_L(t) = \frac{1}{2} p(t) A_I \tag{153}$$

Therefore the total output noise power is

$$\overline{N_L} \equiv \overline{U_L^2(t)} = \frac{1}{2} \overline{p^2(t)} S_{Io}$$

where

$$S_{Io} = A_I^2/2$$

is the average power of the unmodulated carrier at the demodulator input. Equation (138) can then be used to obtain the desired expression for N_L in terms of the total input noise power N_R

$$N_L = \frac{1}{4} G_I N_R \tag{154}$$

where

$$G_I \equiv S_{Io}/S_{Ro} = A_I^2/A_R^2$$

is the overall power gain of the receiver between the receiver input and the demodulator output.

The signal-dependent term is the response*

$$e_L'(t) = \frac{1}{2} a(t) \cos \varphi(t) \tag{155}$$

of the product demodulator to the normalized signal input $e_I(t)$. The response to the total input signal $E_I(t) = A_I e_I(t)$ is then $E_L'(t) = A_I e_L'(t)$, but in interpreting these waveforms, the name "output signal" and the unprimed notations $e_L(t)$ and $E_L(t)$ will be reserved for the portion of the signal-dependent term that is directly proportional to the baseband $e_B(t)$. Whether there is such an output signal or not depends of course on the nature of the modulation specified by $a(t)$ and $\varphi(t)$.

---

*Note that, in terms of the modulation function $m(t) = a(t)e^{j\varphi(t)}$ introduced in Eq. (34), $e_L'(t) = \frac{1}{2} \text{Re } m(t)$.

Referring to Table 1, it is clear that there will be an output signal in the case of the linear modulation methods AM, DSB, and SSB. For these methods, the values of $a(t)$, $\varphi(t)$ given in Table 1, and the resultant expressions for the signal-dependent term $e_L'(t)$ given by Eq. (155) are as follows:

	$a(t)$	$\varphi(t)$	$e_L'(t)$
AM	$1 + Me_B(t)$	0	$\frac{1}{2}\left[1 + Me_B(t)\right]$
DSB	$Me_B(t)$	0	$\frac{1}{2} Me_B(t)$
SSB	$M\sqrt{e_B^2(t) + \check{e}_B^2(t)}$	$\cos^{-1}\dfrac{e_B(t)}{\sqrt{e_B^2(t) + \check{e}_B^2(t)}}$	$\frac{1}{2} Me_B(t)$

Although with AM, the signal-dependent term includes a dc component, the normalized output signal is given by the same expression for all three types of modulation:

$$e_L(t) = E_L(t)/A_I = \frac{1}{2} Me_B(t) \qquad \text{AM, DSB, SSB} \qquad (156)$$

It is interesting to observe how this result comes about. Note especially that, contrary to the impression given by some textbook treatments, the signal-dependent output of a product demodulator is not in general proportional to the instantaneous amplitude of the modulated carrier alone. Instead, it is proportional to the product of the instantaneous amplitude and the cosine of the phase deviation

$\varphi(t)$. With AM and DSB the instantaneous amplitude is proportional to the baseband, but since there is no phase deviation, cos $\varphi(t)$ is unity, and the output signal remains proportional to the baseband. With SSB, on the other hand, the instantaneous amplitude of the carrier is not proportional to the baseband, but rather to a mixture of the baseband with its Hilbert transform. It is only the presence of the cos $\varphi(t)$ factor and its compensatory dependence on both $e_B(t)$ and $\breve{e}_B(t)$ that allows the product demodulator to render a faithful reproduction of the baseband.

The average output signal power may now be determined from Eq. (156)

$$S_L \equiv \overline{E_L^2(t)} = \overline{e_L^2(t)}\, A_I^2 = \frac{1}{2}\frac{M^2}{\Lambda_B} S_{Io} \tag{157}$$

where

$$\Lambda_B \equiv \hat{S}_B/S_B = 1/\overline{e_B^2(t)}$$

is the ratio of peak to average power in the baseband signal.

When the output signal power is expressed in terms of the received input signal power S_R, however, the differences between the methods begin to appear. Thus, using Table 3 to express M^2/Λ_B in terms of S_R, the following relations are obtained:

$$S_L = \begin{cases} \frac{1}{2} G_I (S_R - S_{Ro}) & \text{AM} \\ \frac{1}{2} G_I S_R & \text{DSB} \\ \frac{1}{4} G_I S_R & \text{SSB} \end{cases} \tag{158}$$

where, as before, S_{Ro} is the power of the unmodulated carrier at the receiver input and G_I is the overall power gain of the receiver between its input and the demodulator input. The relation for AM shows that the output signal power is proportional not to the total received signal power, but rather to the amount by which it exceeds the unmodulated carrier component, i.e., to the average power carried by the two sidebands. In the case of DSB and SSB, S_L is proportional to the total received power S_R which is all sideband power since there is no carrier. Thus, for all of the linear modulation methods, the output signal power is directly proportional to total sideband power.

Expressions for the signal-to-noise power ratio K_L at the output of a product demodulator may now be obtained in terms of the signal-to-noise ratio at the receiver input by combining Eqs. (154) and (158)

$$K_L \equiv \frac{S_L}{N_L} = \begin{cases} 2(K_R - K_{Ro}) & \text{AM} \\ 2K_R & \text{DSB} \\ K_R & \text{SSB} \end{cases} \qquad (159)$$

where

$$K_R = S_R/N_R$$

is the ratio of total received signal power to total noise power referred to the receiver input and K_{Ro} is the corresponding ratio for the unmodulated carrier.

Useful alternatives to the AM relation in Eq. (159) may be obtained by eliminating either K_R or K_{Ro}. Thus, dividing the AM expression for S_R in Table 3 by N_R, gives

$$K_{Ro} = \frac{K_R}{1 + M^2/\Lambda_B} \qquad \text{AM} \qquad (160)$$

and so, in terms of the total input signal-to-noise power ratio,

$$K_L = \frac{1}{1 + \Lambda_B/M^2}\, 2K_R \qquad \text{AM} \qquad (161)$$

or, in terms of the ratio of unmodulated carrier power-to-noise,

$$K_L = \frac{M^2}{\Lambda_B}\, 2K_{Ro} \qquad \text{AM} \qquad (162)$$

This last expression, evaluated for the special case of 100 percent modulation ($M = 1$) by a sinusoidal modulating signal ($\Lambda_B = 2$) yields the frequently stated but too often unqualified conclusion that the output and input signal-to-noise ratios are equal in AM.

For comparing the efficiency with which different modulation methods yield a prescribed output signal-to-noise quality, it is the received signal power itself rather than the received signal-to-noise power ratio that is of greater interest. Appropriate relations expressing the output SNR, K_L, for product demodulated AM, DSB, and SSB signals in terms of the average received power S_R and peak received

power $\hat{S}_R$ may be obtained from Eqs. (159), (161), and (162) by substituting the identities

$$K_R \equiv \frac{S_R}{B_R N_o} = \frac{\hat{S}_R}{\Lambda_R B_R N_o}$$

and then using the expressions for B_R and Λ_R given in Table 3.

The results are displayed in Table 5 where it will be noted that, to keep all quantities dimensionless, S_R and $\hat{S}_R$ are divided by $B_B N_o$, the input noise in the baseband bandwidth. Since for a given modulating signal and for receivers of specified noise performance, this normalizing factor is the same for all modulation methods, the tabulated ratios $\frac{S_R/B_B N_o}{K_L}$ and $\frac{\hat{S}_R/B_B N_o}{K_L}$ are directly comparable measures of the received power needed to produce a given output signal quality.

Comment on these results will be deferred until the other types of demodulators have been discussed, but, before leaving the subject of product demodulation, the effect of a constant phase error in the local carrier will be considered. For this purpose, suppose that the local carrier has an initial phase θ_I

$$\cos (\omega_I t + \theta_I)$$

Then after low-pass filtering, the demodulator output is, by analogy with Eq. (144)

$$x_L(t) = x_I(t) \cos (\omega_I t + \theta_I) + \breve{x}_I(t) \sin (\omega_I t + \theta_I)$$

Similarly, in terms of $r(t)$ and $\theta(t)$

$$x_L(t) = \frac{1}{2} r(t) \cos [\theta(t) - \theta_I] \qquad (163)$$

Table 5

INPUT SNR AND INPUT POWER VERSUS OUTPUT SNR

Modulation Method	SNR at Receiver Input K_R/K_L	Average RF Power $\frac{S_R/B_B N_o}{K_L}$	Peak RF Power $\frac{\hat{S}_R/B_B N_o}{K_L}$
AM	$\frac{1}{2}\left(1 + \frac{\Lambda_B}{M^2}\right)$	$1 + \frac{\Lambda_B}{M^2}$	$\left(1 + \frac{1}{M}\right)^2 \cdot 2\Lambda_B$
DSB	$\frac{1}{2}$	1	$2\Lambda_B$
SSB	1	1	$k_1\Lambda_B$ $k_1 = \begin{matrix} 1 \text{ sinusoid} \\ 1.33 \text{ gaussian} \end{matrix}$
PM	$\frac{B_B}{B_R}\Lambda_B/\Phi^2$	Λ_B/Φ^2	$2\Lambda_B/\Phi^2$
FM	$\sim \frac{\Lambda_B}{6(1 + D)D^2}$	$\frac{\Lambda_B}{3D^2}$	$2\frac{\Lambda_B}{3D^2}$
SSBFM	$\sim \frac{\Lambda_B}{6(1 + D)D^2}$	$\sim \frac{2\Lambda_B}{9D^2}$	$k_2\Lambda_B$

and, in terms of $a(t)$, $\varphi(t)$, $v(t)$, and $\psi(t)$

$$x_L(t) = \frac{1}{2}\left\{a(t)\cos\left[\varphi(t) - \theta_I\right] + v(t)\cos\left[\psi(t) - \theta_I\right]\right\}$$

Comparing this expression with Eq. (151), it is seen that the effect of the phase error is to shift the phase of both signal- and noise-dependent components of the demodulator output by θ_I. The output noise is now

$$u_L(t) = \frac{1}{2}\, v(t)\cos\left[\psi(t) - \theta_I\right]$$

and, for the linear modulation methods, the signal-dependent term $\frac{1}{2}\, a(t)\cos\left[\varphi(t) - \theta_I\right]$ takes on the following values:

$$\frac{1}{2}\cos\theta_I + \cos\theta_I \cdot \frac{1}{2}\, Me_B(t) \qquad \text{AM}$$

$$\cos\theta_I \cdot \frac{1}{2}\, Me_B(t) \qquad \text{DSB}$$

$$\pm \sin\theta_I \cdot \frac{1}{2}\, \check{Me}_B(t) + \cos\theta_I \cdot \frac{1}{2}\, Me_B(t) \qquad \text{SSB}$$

where the upper sign in the SSB expression corresponds to use of the upper sideband and the lower sign to the lower sideband. Thus, the output signal--i.e., the part of this term that is proportional to the baseband--is again the same for all three modulation methods

$$e_L(t) = E_L(t)/A_I = \cos\theta_I \cdot \frac{1}{2}\, Me_B(t)$$

but is now reduced in amplitude by the factor $\cos\theta_I$.

With AM, the signal-dependent term includes a dc term as before, but both terms are reduced in magnitude by the factor $\cos\theta_I$. With SSB on the other hand, there is an entirely new term which varies with time in proportion to the Hilbert transform of the baseband. For small

phase error, the sin θ_I factor in this term makes it negligible compared with $e_L(t)$, but for larger values of θ_I it can have a magnitude equal to or greatly exceeding $e_L(t)$. Moreover, the new term cannot be separated from $e_L(t)$. Together, they produce a resultant that is the same as would be obtained by passing the baseband through a linear filter which shifts the phase of all its Fourier components by θ_I radians.

The conclusion is that for AM and DSB, a phase error in the local carrier merely reduces the magnitude of the output signal, whereas with SSB, it produces a phase distortion. Since this type of distortion can seriously deform pulse shapes and degrade video basebands, correct local carrier phase is essential to the use of SSB for transmitting such signals. However, voice signals are relatively insensitive to phase distortion and can tolerate not only phase errors in the local carrier but also frequency errors of up to a few tens of cycles per second.

To determine the effect of a constant phase error θ_I on the output signal-to-noise ratio K_L in a product demodulator, it is only necessary to note that, since $\psi(t)$ is uniformly distributed, the output noise power is unaffected

$$N_L = \overline{U_L^2(t)} = \frac{1}{4} A_I^2 \overline{v^2(t) \cos^2 [\psi(t) - \theta_I]} = \frac{1}{4} G_I N_R$$

whereas the signal power is reduced by the factor $\cos^2 \theta_I$ from the values given in Eq. (158). Hence, all of the expressions for K_L must be reduced by this same factor. In particular, when the local carrier phase error is an odd multiple of $\pi/2$ the output signal and the output signal-to-noise ratio vanish regardless of the input signal strength.

With SSB, however, a distorted version of the baseband remains and may contain as much power as would the baseband itself in the absence of a phase error.

Envelope Demodulator

An ideal envelope demodulator is one which, given the band-limited input waveform of Eq. (139)

$$x_I(t) = r(t) \cos\left[\omega_I t + \theta(t)\right]$$

yields an output proportional to the envelope r(t), providing that r(t) and the phase deviation $\theta(t)$ are defined in terms of the analytic signal representation $\underline{x}_I(t)$ of $x_I(t)$ by Eqs. (145) and (146).

To verify this, consider one model of an ideal envelope detector--a square-law full-wave rectifier followed by an ideal low-pass filter and a circuit which takes the square root of the filter output. If $x_I(t)$ is written in terms of its analytic signal representation as shown in Eq. (142), the rectifier output is

$$x_I^2(t) = \frac{1}{4}\left[\underline{x}_I^2(t) + 2\underline{x}_I(t)\,\underline{x}_I^*(t) + \underline{x}_I^{*\,2}(t)\right]$$

Let the frequency limits on the band occupied by $x_I(t)$ be f_1 and $f_2 = f_1 + B_R$. Then, since the spectrum of a product of two time functions is the convolution of their individual spectra, the spectrum of the first term is confined to the range $-2f_2 < f < 2f_1$, the second term to $|f| < f_2 - f_1$, and the third to $2f_1 < f < 2f_2$. Hence the output of an ideal low-pass filter with cutoff frequency B_R is just the second term

$$\frac{1}{2}\,\underline{x}_I(t)\,\underline{x}_I^*(t) \equiv \frac{1}{2}\,|\underline{x}_I(t)|^2$$

and the normalized demodulator output $x_L(t)$ is its square-root,

$$x_L(t) = \frac{1}{\sqrt{2}} \left|\underline{x}_I(t)\right| = \frac{1}{\sqrt{2}} r(t) \qquad (164)$$

Other idealized models of an envelope detector yield the same dependence on $r(t)$ although with different constants of proportionality, and, in some cases, additional restrictions on the form of the input $x_I(t)$. In practical envelope detectors, the output is an approximation to $r(t)$ which improves as $x_I(t)$ becomes more narrow-band, i.e., as B_R/f_I decreases.

As with the product demodulator, the important point is that, even though $x_I(t)$ might have been produced by a combination of amplitude ratio $r'(t)$ and phase deviation $\theta'(t)$ different from the combination $r(t)$, $\theta(t)$ defined by the analytic signal, the output of the envelope detector is $r(t)$, and not $|r'(t)|$. This is the ultimate justification for restricting the term "envelope" to the modulus of the analytic signal representation.

To identify the output signal and noise components in the output $x_L(t)$, substitute Eq. (140) into Eq. (164) to obtain

$$x_L(t) = \frac{1}{\sqrt{2}} \left|a(t)\right| \cdot \rho(t)$$

where

$$\rho(t) \equiv \sqrt{1 + 2\,\frac{v(t)}{a(t)} \cos\left[\psi(t) - \varphi(t)\right] + \left[\frac{v(t)}{a(t)}\right]^2}$$

Several things should be noted about this expression. First, unlike the product demodulator output, the envelope demodulator output does not contain a separate term which depends only on the modulation

of the received signal as given by $a(t)$ and $\varphi(t)$. Second, the instantaneous amplitude ratio $a(t)$ appears in absolute value. Third, the time-dependent factor $\rho(t)$, which may be regarded as an additional amplitude modulation due to noise, depends on the input signal as well as the noise. Finally, the magnitude of $\rho(t)$ approaches unity for large values of the input signal-to-noise ratio, $K_R = \overline{a^2(t)}/\overline{v^2(t)}$. The conclusion is that the envelope demodulator can yield a faithful reproduction of the imposed amplitude ratio $a(t)$ only under the conditions

Large input SNR	$K_R \gg 1$
Nonnegative amplitude ratio	$a(t) \geq 0$

When these conditions are met, the output $x_L(t)$ can be written

$$x_L(t) = e_L'(t) + u_L(t)$$

where a signal-dependent term

$$e_L'(t) \equiv \frac{1}{\sqrt{2}}\, a(t), \quad a(t) \geq 0 \tag{165}$$

has arbitrarily been separated from a noise term

$$u_L(t) = \epsilon(t)\, e_L'(t) \tag{166}$$

containing the time-dependent factor

$$\epsilon(t) \equiv \rho(t) - 1 \cong \frac{v(t)}{a(t)} \cos\left[\psi(t) - \varphi(t)\right], \quad K_R >> 1 \tag{167}$$

Substituting Eqs. (165) and (167) into Eq. (166) gives

$$u_L(t) \cong \frac{1}{\sqrt{2}}\, v(t) \cos\left[\psi(t) - \varphi(t)\right], \quad K_R >> 1 \tag{168}$$

For AM with modulation index M, the amplitude ratio and phase deviation are respectively $a(t) = 1 + Me_B(t)$, and $\varphi(t) = 0$. Again, the

normalized output signal $e_L(t)$ is identified with the term in the signal-dependent component $e'_L(t)$ that is proportional to the baseband $e_B(t)$.

$$e_L(t) \equiv E_L(t)/A_I \equiv e'_L(t) - \frac{1}{\sqrt{2}} = \frac{1}{\sqrt{2}} Me_B(t), \qquad M \leq 1$$

Similarly, substituting for $a(t)$ and $\varphi(t)$ in Eq. (168), the normalized output noise with AM is

$$u_L(t) \equiv U_L(t)/A_I \cong \frac{1}{\sqrt{2}} v(t) \cos \psi(t) = \frac{1}{\sqrt{2}} p(t), \qquad \begin{matrix}\text{AM with} \\ K_R >> 1\end{matrix}$$

Comparison with Eqs. (156) and (152) shows that for the assumed conditions on input signal-to-noise ratio and modulation index, the signal and noise outputs of an envelope detector differ from those of a product detector only by a constant factor. Consequently the signal-to-noise relations shown for AM in Table 5 apply equally well to envelope demodulation when $K_R >> 1$ and $M \leq 1$.

For DSB and SSB, the amplitude ratio does not meet the condition $a(t) \geq 0$. With DSB, however, insertion of a properly phased local carrier prior to demodulation effectively converts the DSB waveform into the AM type just considered. Thus if the input signal-to-noise ratio is high, and if the local carrier amplitude is no less than the peak instantaneous amplitude of the DSB signal (to keep the effective modulation index below 100 percent), an envelope demodulator will yield virtually the same outputs as a product demodulator. An envelope demodulator can be used in a similar fashion to demodulate a SSB waveform, but in addition to a high input signal-to-noise ratio and accurate local carrier phase, the amplitude of the local carrier must greatly exceed the peak input signal level if output signal distortion is to be avoided.[10]

Compared with product demodulation, these methods of using an envelope detector with DSB and SSB modulated carriers have found little favor in practice. They require the generation of a local carrier no less than does a product detector, offer no significant reduction in circuit complexity, and yield a lower output signal quality when the input signal level drops. Ingenious methods have been proposed to avoid the need for synchronous carrier reinsertion in the envelope detection of SSB signals,[16] but the cost in circuit complexity for the necessary signal processing is often high.

It has been emphasized that for an envelope demodulator (but not for a product demodulator) the expressions for output SNR given in Table 5 are valid only when the input SNR K_R is large compared with unity. To be more explicit, i.e., to determine just how large K_R must be, would require an analysis of how the input and output SNRs are related as K_R decreases toward unity (equal signal and noise inputs). Such an analysis is difficult, as may be appreciated from an inspection of Eq. (166), the general expression for the noise-dependent component of the demodulator output, and will not be given here. Suffice it to say that with an envelope detector, there is a threshold value of input SNR below which the output signal quality (to the extent that an output "signal" is even identifiable) falls off more rapidly than in direct proportion to the input SNR. This AM threshold is not sharply defined, but since it occurs well below the values of K_R necessary for commercially acceptable output signal quality, its existence is of little practical importance.

Phase Demodulator

With a narrow-band input as in Eq. (139)

$$x_I(t) = r(t)\cos\left[\omega_I t + \theta(t)\right]$$

an ideal phase demodulator yields an output proportional to the phase deviation $\theta(t)$, defined as before by

$$\omega_I t + \theta(t) = \arg \underline{x}_I(t)$$

where $\underline{x}_I(t)$ is the analytic signal representation of the input $x_I(t)$. One prototype of such a detector consists of an ideal limiter followed by a product demodulator and a baseband filter. It operates as follows.

The limiter first converts $x_I(t)$ into an infinitely clipped or squared-off wave having the same zero crossings

$$\operatorname{sgn}\left[x_I(t)\right] = \begin{cases} 1 & x_I > 0 \\ 0 & x_I = 0 \\ -1 & x_I < 0 \end{cases}$$

and then passes it through a band-pass filter centered on ω_I to retain the constant amplitude "fundamental"

$$\cos\left[\omega_I t + \theta(t)\right] \tag{169}$$

The effect of the limiter is thus to remove the time variation from the envelope of the composite signal-plus-noise waveform $x_I(t)$ without affecting its phase deviation.

The product demodulator, which differs from that used with the linear modulation methods only in that its local carrier is in phase

quadrature with the unmodulated carrier, operates on this waveform to yield an output given by Eq. (163) with envelope $r(t) = 1$ and "phase error" $\theta_I = \pi/2$:

$$X_L'(t) = \tfrac{1}{2} \sin \theta(t)$$

Upper case symbols are used for this and subsequent output waveforms because, due to the effect of the limiter, they represent the total phase demodulator output regardless of whether the input is normalized or not. If the peak value of the phase deviation is small compared with a radian, as will henceforth be assumed, this output is to a very good approximation just

$$X_L'(t) \cong \tfrac{1}{2}\, \theta(t) \qquad \theta(t) \ll 1 \tag{170}$$

This condition on $\theta(t)$ is not as restrictive as it appears because, with the aid of a "phase-lock loop" feedback circuit,[17,18] the net phase deviation at the demodulator input can be kept small without similarly restricting the phase deviation of the received signal.

In any case, the output X_L' occupies a bandwidth $B_R/2$ (for double-sidebanded carriers) or B_R (for single-sidebanded carriers) as determined by the output filter of the product demodulator, and it is then applied to an ideal baseband filter, i.e., a rectangular low-pass filter with cutoff frequency equal to the baseband bandwidth B_B. To determine the effect of this baseband filter, it is necessary to separate $X_L'(t)$ into signal-dependent and noise-dependent components.

For this purpose, a more useful expression for $\theta(t)$ than Eq. (141) is (see Fig. 9)

$$\theta(t) = \varphi(t) + \eta(t) \tag{171}$$

where $\varphi(t)$ is the phase deviation of the input signal $e_I(t)$, and

$$\eta(t) \equiv \tan^{-1} \frac{v(t) \sin[\psi(t) - \varphi(t)]}{a(t) + v(t) \cos[\psi(t) - \varphi(t)]} \tag{172}$$

is the incremental phase deviation at the product demodulator output due to the input noise $u_I(t)$.

Substituting Eq. (171) into Eq. (170) yields the desired expression

$$X_L'(t) \equiv E_L'(t) + U_L'(t) \qquad \theta(t) \ll 1 \tag{173}$$

where the purely signal-dependent term

$$E_L'(t) \equiv \tfrac{1}{2} \varphi(t) \tag{174}$$

has been separated from a noise-dependent term

$$U_L'(t) \equiv \tfrac{1}{2} \eta(t) \tag{175}$$

which, as in the case of the envelope detector, is a function of the input signal modulation as well as the noise. However, when the input signal-to-noise ratio $K_R = \overline{a^2(t)}/\overline{v^2(t)}$ is large, $\eta(t)$ is small, and to a good approximation

$$\eta(t) \simeq \frac{v(t)}{a(t)} \sin\left[\psi(t) - \varphi(t)\right] \qquad K_R \gg 1 \tag{176}$$

Moreover, so far as the effect of the baseband filter on noise power is concerned, $\eta(t)$ is relatively insensitive to the nature of the input signal modulation specified by $a(t)$ and $\varphi(t)$. In particular, an unmodulated carrier, ($a(t) = 1$, $\varphi(t) = 0$), can be assumed and Eqs. (175) and (176)

combined to yield for the noise-dependent component

$$U_L'(t) = \tfrac{1}{2}\ \eta(t) \simeq \tfrac{1}{2}\ v(t) \sin \psi(t) \equiv \tfrac{1}{2}\ q(t) \qquad K_R \gg 1 \qquad (177)$$

where $q(t)$ is the quadrature component of the input noise waveform defined in Eq. (129). Note the similarity of this result to the expression for the normalized noise output of a product demodulator given in Eq. (152).

The output of the phase demodulator in this case is

$$X_L(t) = E_L(t) + U_L(t)$$

where $E_L(t)$ and $U_L(t)$ are respectively the responses of the baseband filter to the inputs $E_L'(t)$ and $U_L'(t)$ specified by Eqs. (174) and (175). It is not necessary to have an explicit expression for the output noise waveform $U_L(t)$; computation of the output noise power requires only a knowledge of its power spectrum. Since the baseband filter is rectangular and lossless, it follows that

$$W_{U_L}(f) = W_{U_L'}(f) = \tfrac{1}{4}\ W_q(f) = \begin{cases} (1/8)\ N_o/S_{Ro} & |f| \le B_B \\ 0 & |f| \ge B_B \end{cases} \qquad (178)$$

where the last pair of equations follows from Eqs. (133) and (138). It follows that the output noise power is

$$N_L = \int_{-\infty}^{\infty} W_{U_L}(f)df = 1/4\ N_o B_B/S_{Ro} \qquad (179)$$

Recalling that $N_o B_R = N_R$ this may also be written

$$N_L = \tfrac{1}{4} \frac{B_B}{B_R} \frac{1}{K_{Ro}} \tag{180}$$

where $K_{Ro} \equiv S_{Ro}/N_R$ is the unmodulated carrier-to-noise power ratio at the receiver input. The fact that the noise power at the output of a phase detector is inversely proportional both to the RF bandwidth and the carrier-to-noise ratio is in marked contrast to the behavior of the detectors appropriate to the linear modulation methods.

Strictly speaking, a phase demodulator yields an output signal proportional to the baseband only when PM is used. In this case, as in Table 1, $\varphi(t) = \Phi e_B(t)$, where Φ is the peak phase deviation. Hence the signal component $E_L'(t) = \frac{1}{2}\varphi(t) = \frac{1}{2}\Phi e_B(t)$ is passed without change by the baseband filter,

$$E_L(t) = E_L'(t) = \tfrac{1}{2}\Phi e_B(t) \qquad \text{PM}$$

so the output signal power is

$$S_L = \overline{E_L^2(t)} = \tfrac{1}{4}\,\Phi^2\,\overline{e_B^2(t)} = \tfrac{1}{4}\,\frac{\Phi^2}{\Lambda_B^2} \qquad \text{PM} \tag{181}$$

Note that the output signal power is independent of the input signal power. Again this is in contrast to the linear modulation methods, but is not unexpected since, as noted in Table 3, the input signal power does not depend on whether the carrier is modulated or not, i.e.,

$$S_R = S_{Ro} \qquad \text{PM}$$

Combining this fact with Eqs. (180) and (181) leads to the relation between input and output signal-to-noise ratios for PM:

$$K_L \equiv \frac{S_L}{N_L} = \frac{\Phi^2}{2} \frac{B_R}{\Lambda_B} \frac{}{B_B} K_{Ro} \qquad \text{PM} \qquad (182)$$

This expression has been entered in Table 5 along with the values for the other ratios which may be derived from it with the aid of Table 3.

A phase demodulator can also be used to demodulate an FM or SSBFM waveform by merely adding an ideal differentiator just prior to the baseband filter. For both types of modulation, the phase deviation is

$$\varphi(t) = \Omega \int e_B(t)\, dt$$

where Ω is the peak frequency deviation. Thus, by Eq. (174), the signal input to the differentiator is

$$E_L'(t) = \tfrac{1}{2}\, \Omega \int e_B(t) dt$$

Its output is of course proportional to the baseband signal itself and so is passed without change by the baseband filter to yield

$$E_L(t) = \dot{E}_L'(t) = \tfrac{1}{2}\, \Omega\, e_B(t)$$

Hence the output signal power is

$$S_L \equiv \overline{E_L^2(t)} = \tfrac{1}{4}\, \Omega^2 \overline{e_B^2(t)} = \tfrac{1}{4} \frac{\Omega^2}{\Lambda_B} \qquad \text{FM, SSBFM} \qquad (183)$$

In the case of the noise, the differentiator converts the product detector output given by Eq. (177)

$$U_L'(t) = \tfrac{1}{2}\, q(t) \tag{184}$$

into a waveform $\dot{U}_L'(t)$ whose power spectrum is simply related to that of $U_L'(t)$. Thus, using Eqs. (62b), (184), and (178),

$$W_{\dot{U}_L'}(f) = \omega^2 W_{U_L'}(f) = \tfrac{1}{4}\,\omega^2 W_q(f) = \begin{cases} (1/8)\omega^2 N_o/S_{Ro} & |f| < B_R \\ 0 & |f| > B_R \end{cases}$$

With this input, the average noise power at the output of the baseband filter is thus

$$N_L \equiv \int_{-B_B}^{B_B} W_{\dot{U}_L'}(f)\, dt = \frac{\pi^2 N_o}{S_{Ro}} \int_0^{B_B} f^2\, df = \frac{1}{4}\,\frac{4\pi^2 B_B^3}{3B_R K_{Ro}} \tag{185}$$

where as before, K_{Ro} is the unmodulated carrier-to-noise ratio at the receiver input. Combining Eqs. (184) and (185) gives for the output signal-to-noise ratio in FM and SSBFM,

$$K_L \equiv \frac{S_L}{N_L} = \frac{3D^2}{\Lambda_B}\,\frac{B_R}{B_B}\,K_{Ro} \qquad \begin{matrix} \text{FM, SSBFM} \\ K_{Ro} >> 1 \end{matrix} \tag{186}$$

where

$$D = \frac{\Omega}{2\pi B_B} \tag{187}$$

is the deviation ratio or FM modulation index. For FM, but not for SSBFM, the unmodulated carrier power and the modulated carrier power are equal and K_{Ro} can be replaced by K_R. Not surprisingly, the results for FM and SSBFM using a phase detector with differentiator are identical to those for the frequency demodulator discussed on the next page.

Just as with an envelope detector, the expressions for the SNR at the output of a phase detector are valid only for input signal-to-noise ratios that exceed a certain threshold value. Below this value of K_R, the statistical nature of the output noise changes and the output SNR deteriorates more rapidly than in direct proportion to K_R. In contrast to envelope detection, however, the threshold phenomenon with a phase detector is of considerable practical importance because it can occur when the output SNR is quite high. Without the threshold restriction, the SNR equations imply that a specified high output SNR can be achieved with PM or FM using as little input signal power as desired, merely by choosing a sufficiently large modulation index. As it is, increasing the PM or FM modulation index permits the input signal power to be reduced while maintaining the output SNR fixed, but the increase in modulation index is accompanied by an increase in the RF bandwidth and hence in the input noise. With the signal decreasing and the noise increasing, the input SNR soon reaches its threshold value; the corresponding modulation index is then the optimum for the combination of baseband signal, output SNR, and receiver noise performance in question.

Expressions for the threshold in the case of FM signals together with a brief discussion of a feedback technique for threshold reduction will be found in the following treatment of frequency demodulators.

Frequency Demodulator

An ideal frequency demodulator or "discriminator" converts a narrow-band signal

$$x_I(t) = r(t) \cos\left[\omega_I t + \theta(t)\right]$$

into an output whose ac component is proportional to the instantaneous frequency $\dot{\theta}(t)$, where $\omega_I t + \theta(t)$ is again the argument of the analytic signal representation of $x_I(t)$. An idealized model of such a detector consists of a limiter followed by a differentiator, an envelope detector, and a baseband filter.* It operates as follows.

As in the phase demodulator, the limiter removes the amplitude variations from $x_I(t)$ to yield the output given by Eq. (169). This is then converted by the differentiator into the signal

$$\left[\omega_I + \dot{\theta}(t)\right] \cos\left[\omega_I t + \theta(t) + \pi/2\right]$$

which is narrow band so long as the bandwidth of $\dot{\theta}(t)$ is small compared with f_I. Assuming this to be the case, the envelope detector output is given by Eq. (164) with $r(t)$ replaced by $\omega_I + \dot{\theta}(t)$:

$$X_L'(t) = \frac{1}{\sqrt{2}}\left[\omega_I + \dot{\theta}(t)\right] \tag{188}$$

As with the other detectors, the effect of the baseband filter on this waveform may be found by separating it into signal- and noise-dependent components. For this purpose differentiate Eq. (171) to obtain

$$\dot{\theta}(t) = \dot{\varphi}(t) + \dot{\eta}(t) \tag{189}$$

where $\dot{\varphi}(t)$ is the time derivative of the phase deviation, or simply the "phase rate," of the input signal $e_I(t)$, and $\dot{\eta}(t)$ is the time derivative

*The combination of differentiator, detector, and filter, or its functional equivalent, is often called a "discriminator."

of the incremental phase deviation due to noise given by Eq. (172).

For output noise calculations when the input SNR is high, however, $\eta(t)$ is well approximated by Eq. (177), which makes

$$\dot{\eta}(t) \cong \dot{q}(t) \qquad K_R \gg 1 \tag{190}$$

Combining Eqs. (188), (189), and (190) yields the desired expression for the input to the baseband filter

$$X_L'(t) = E_L'(t) + U_L'(t) \tag{191}$$

where, except for a dc term, the signal component is

$$E_L'(t) = \frac{1}{\sqrt{2}} \dot{\varphi}(t) \tag{192}$$

and the noise component is

$$U_L'(t) = \frac{1}{\sqrt{2}} \dot{q}(t) \tag{193}$$

The output of the frequency demodulator may be written in the same fashion

$$X_L(t) = E_L(t) + U_L(t)$$

where $E_L(t)$ and $U_L(t)$ are the responses of the baseband filter to $E_L'(t)$ and $U_L'(t)$, respectively.

To find the output noise power, note that below the cutoff frequency of the baseband filter the power spectrum of $U_L(t)$ is the same as that of $U_L'(t)$. But, by virtue of Eq. (193) and Eq. (62a), $W_{U_L'}(f)$ is readily expressed in terms of $W_q(f)$, the power spectrum of the quadrature component of the input noise,

$$W_{U_L}(f) = W_{U_L'}(f) = \tfrac{1}{2}W_{\dot{q}}(f) = \tfrac{1}{2}\omega^2 W_q(f) \qquad |f| \le B_B$$

Substituting from Eq. (178), the output noise spectrum is thus

$$W_{U_L}(f) = \begin{cases} \frac{1}{4}\, 4\pi^2 f^2\, N_o/S_{Ro} = \frac{1}{4}\,\frac{4\pi^2 f^2}{B_R K_{Ro}}, & |f| \le B_B \\ 0, & |f| > B_B \end{cases} \qquad (194)$$

where $K_{Ro} = S_{Ro}/N_R = S_{Ro}/B_R N_o$ is the unmodulated carrier-to-noise ratio. Hence the output noise power is

$$N_L = \int_{-\infty}^{\infty} W_{U_L}(f)\, df = \tfrac{1}{2}\,\frac{4\pi^2 B_B^3}{3B_R K_{Ro}} \qquad (195)$$

which differs only by a constant factor from Eq. (185) for the noise output of a phase demodulator with an added differentiator.

The output signal $E_L(t)$ depends of course on the modulation method; in particular it depends on how the imposed phase deviation is related to the baseband. For FM and SSBFM, $\varphi(t) = \Omega \int e_B(t)\, dt$ so, by Eq. (192), the signal input to the baseband filter is proportional to the baseband signal and is passed without change:

$$E_L(t) = E_L'(t) = \frac{1}{\sqrt{2}}\,\dot{\varphi}(t) = \frac{1}{\sqrt{2}}\,\Omega\, e_B(t)$$

Hence the output signal power is

$$S_L = \tfrac{1}{2}\,\Omega^2\,\overline{e_B^2(t)} = \tfrac{1}{2}\,\Omega^2/\Lambda_B \qquad (196)$$

and, combining this with Eq. (195), the output SNR is identical to Eq. (186) for the phase detector with differentiator:

$$K_L \equiv \frac{S_L}{N_L} = \frac{3D^2}{\Lambda_B}\frac{B_R}{B_B} K_{Ro} \qquad \begin{matrix}\text{FM, SSBFM} \\ K_{Ro} >> 1\end{matrix} \qquad (197a)$$

It hardly needs to be added that a frequency demodulator with an integrator ahead of the baseband filter will recover the baseband from a PM signal just as effectively as a phase demodulator. As before, the SNRs are related as in Eq. (182).

An alternative and often quite useful version of Eq. (197a) may be obtained by first noting that

$$\frac{D^2}{\Lambda_B} = D_{rms}^2$$

where D_{rms} is the rms deviation ratio defined as the ratio of the rms value of the instantaneous frequency deviation to the highest baseband frequency:

$$D_{rms} \equiv \frac{\sigma_{\dot{\varphi}}}{2\pi f_m}$$

Making this substitution, and using the identity

$$\frac{B_R}{B_B} K_{Ro} \equiv \frac{S_{Ro}}{B_B N_o}$$

Eq. (197a) may be written in the somewhat simpler forms

$$K_L = 3\frac{D^2}{\Lambda_B}\frac{S_{Ro}}{B_B N_o} = 3 D_{rms}^2 \frac{S_{Ro}}{B_B N_o} \qquad \begin{matrix}\text{FM, SSBFM} \\ S_{Ro} >> B_R N_o\end{matrix} \qquad (197b)$$

Note that the SNR on the right-hand side of this equation is just the ratio of average carrier power to the noise at the receiver input contained in the baseband bandwidth B_B.

The most frequently cited version of Eq. (197a), however, is actually the special case obtained by assuming that the modulating baseband is a single sinusoid. In this case, $\Lambda_B = 2$ and Eq. (197a) reduces to

$$K_L = 3D^2 \frac{S_{Ro}}{2B_B N_o} \qquad \text{FM by a sinusoid} \quad S_{Ro} >> B_R N_o \qquad (197c)$$

Here the SNR on the right side of the equation is the ratio of carrier power to noise power in twice the baseband bandwidth.

The corresponding special case for 100 percent modulated AM may be obtained by substituting $M = 1$ and $\Lambda_B = 2$ in Eq. (162):

$$K_L = \frac{S_{Ro}}{2B_B N_o} \qquad \text{100 percent AM by sinusoid}$$

The fact that the output SNR is higher for FM by the factor $3D^2$ is usually expressed by saying that the FM "improvement factor" (relative to AM) is $3D^2$. Writing Eqs. (162) and (197a) in terms of S_{Ro} and B_B, it is seen that this statement is not restricted to sinusoidal modulation but holds for all basebands with the following important restrictions:

1. The AM modulation index M is equal to unity (100 percent modulation).
2. The FM modulation index D is sufficiently small that the input SNR, $K_{Ro} \equiv \frac{S_{Ro}}{B_R N_o}$, exceeds its threshold value (see below).

It is also important to recognize that this comparison of output SNRs with AM and FM is based on the assumption of equal carrier powers S_{Ro} in the absence of modulation, rather than on equal average modulated carrier powers S_R. Referring to Table 3 it is seen that with $M = 1$,

$$S_R = \begin{cases} (1 + \Lambda_B^{-1})\ S_{Ro} & \text{AM} \\ S_{Ro} & \text{FM} \end{cases}$$

Thus, on the basis of equal S_R, the FM improvement factor is

$$\left(1 + \frac{1}{\Lambda_B}\right) 3D^2$$

For a single speech channel ($\Lambda_B \sim 60$) or a multiplex of many speech channels ($\Lambda_B \sim 10$) this represents only a slight added advantage for FM, but with sine-wave modulation the improvement is $4.5\ D^2$ rather than $3D^2$.

This completes the derivation of signal-to-noise relations for the conventional frequency demodulator. It remains to elaborate briefly on the range of input signal-to-noise ratios for which the SNR expressions is valid, i.e., to specify the threshold value of K_R.

Qualitatively, the threshold behavior of phase and frequency demodulators can be understood with the aid of the vector diagram in Fig. 10. This differs from Fig. 9 only in the relative magnitudes of the signal and noise vectors, and it can be seen that Eq. (141) for the phase deviation $\theta(t)$ of the signal-plus-noise input $x_I(t)$ still holds. But now, with the signal and noise vectors comparable in magnitude on the average, there are frequent occasions when small and slow changes in their respective phase deviations $\varphi(t)$ and $\psi(t)$ can cause large and rapid changes in $\theta(t)$. Since the output noise $U_L'(t)$ is proportional to $\eta(t) \equiv \theta(t) - \varphi(t)$ for a phase detector, and to $\dot{\eta}(t)$ for a frequency detector, it too will display sudden large amplitude changes. The occurrence of these impulses is but one of a

Unmodulated carrier waveform	$\text{Re } e^{j\omega_I t} = \cos \omega_I t$
Signal waveform	$e_I(t) = \text{Re } \underline{e}_I(t) = a(t) \cos [\omega_I t + \varphi(t)]$
Noise waveform	$u_I(t) = \text{Re } \underline{u}_I(t) = v(t) \cos [\omega_I t + \psi(t)]$
Signal plus noise waveform	$x_I(t) = \text{Re } \underline{x}_I(t) = r(t) \cos [\omega_I t + \theta(t)]$

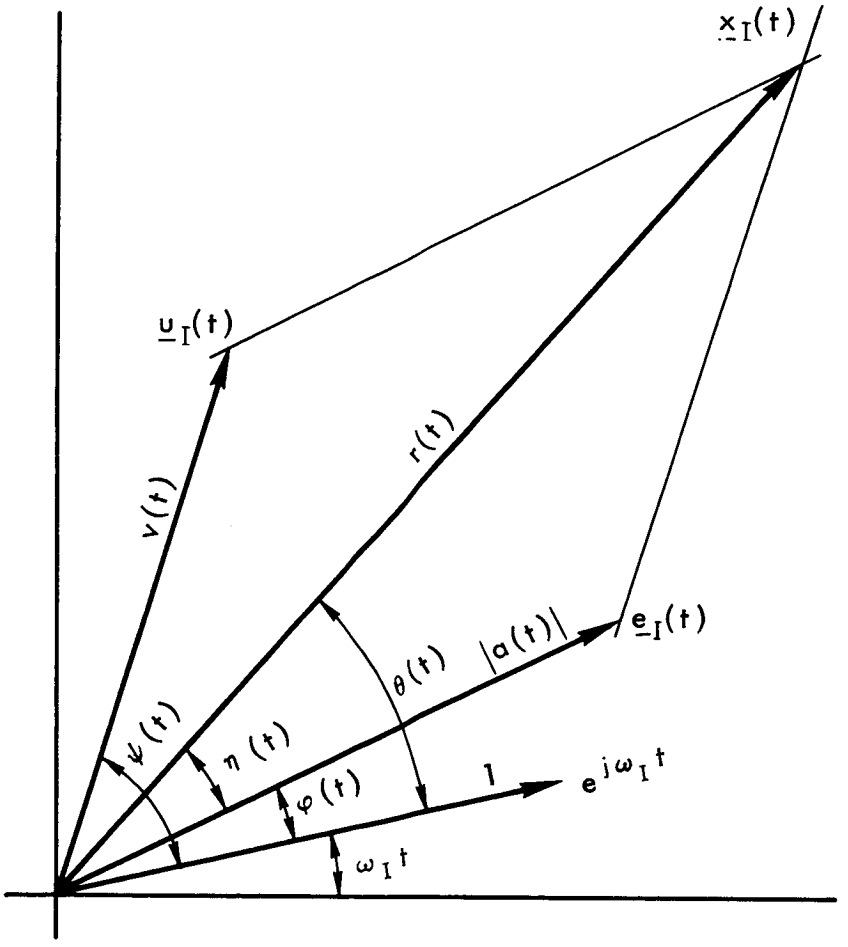


Fig.10— Rotating vector representation of normalized waveforms at the demodulator input when the input SNR $K_R \cong 1$

number of complex changes that affect the statistical characteristics of the output noise as the input carrier-to-noise ratio decreases. The overall result is that the power spectrum of the output noise becomes more uniform; in particular, it no longer vanishes as $f \to 0$.

To show this behavior quantitatively, the output noise power spectrum $W'_{U_L}(f)$ of a frequency demodulator* as calculated by Stumpers[17] is given for a number of different values of K_R in Fig. 11a. The gradual change from the parabolic spectrum given by Eq. (194) and characteristic of large values of K_R to the more uniform spectra appropriate to small values of K_R is clearly evident.

Given these curves for the noise power spectra at the input to the baseband filter, the output noise power is obtained for various values of K_R and B_R by numerical evaluation of the integral

$$N_L = \int_{-B_B}^{B_B} W_{U_L'}(f)\, df$$

When the values of noise power thus obtained are combined with values of output signal power obtained from Eq. (196), the signal-to-noise curves of Fig. 11b can be plotted. Here, the output SNR, K_L, is shown as a function of the carrier-to-noise ratio K_R for various values of the ratio of RF bandwidth to baseband bandwidth. It will be recalled that this ratio can be expressed in terms of the FM modulation index D, using Carson's rule, Eq. (125):

---

*The corresponding curves for a phase detector are not shown since they may readily be obtained from Fig. 11a by dividing $W'_{U_L}(f)$ by ω^2 at each frequency $f = \omega/2\pi$.

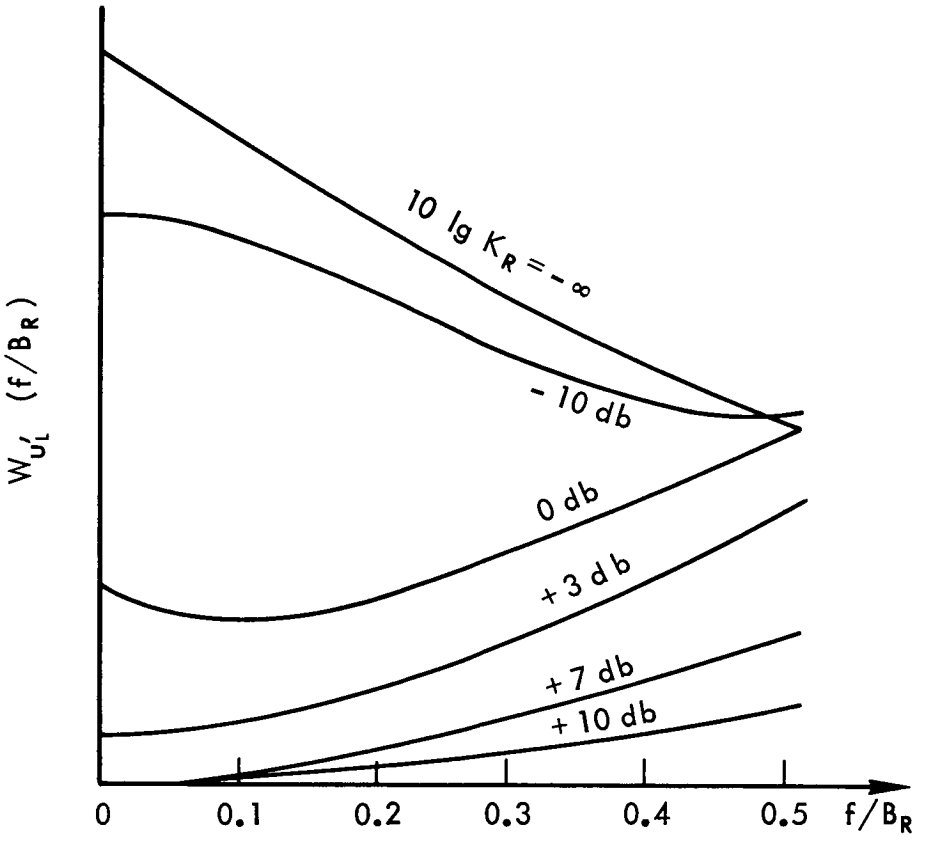


Fig.11a—Output noise power spectrum of frequency demodulator for various values of input SNR, K_R

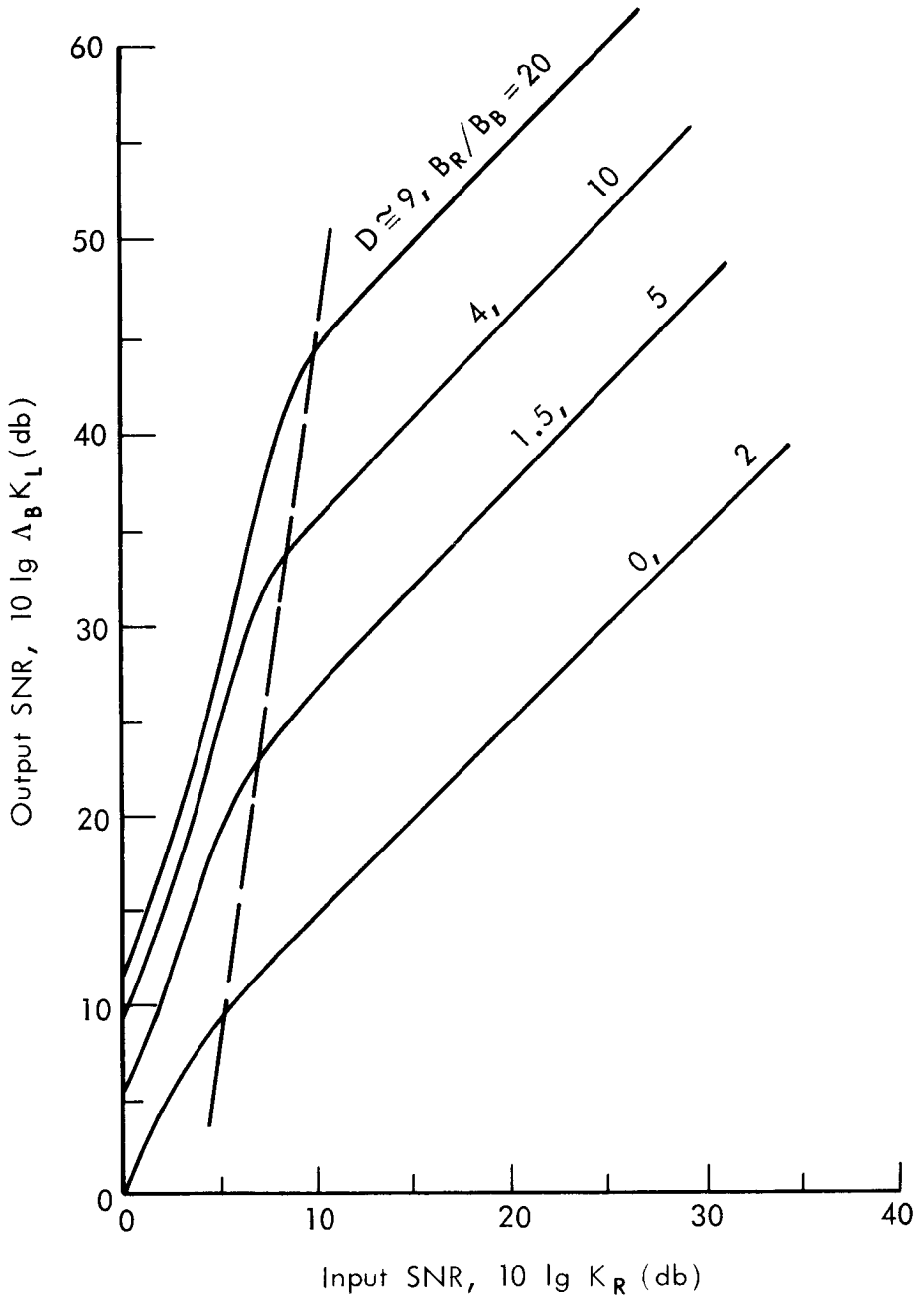


Fig. 11b—Output SNR versus input SNR for frequency demodulator and various values of FM modulation index D

$$\frac{B_R}{B_B} \cong 2(D + 1) \tag{198}$$

The curves clearly display the behavior already described qualitatively: for each value of B_R/B_B, there is a threshold value of K_R above which K_L is directly proportional to K_{Ro} as in Eq. (197), whereas below the threshold, K_L falls off much more rapidly. The choice of an exact value K_{Rt} for the threshold is somewhat arbitrary, but if it is defined as the value of K_R for which K_L lies 0.5 db below that computed from Eq. (197), the values corresponding to the dashed line in Fig. 11b are obtained.* Note that when defined in this manner, the threshold SNR increases with increasing deviation ratio in the manner shown in Fig. 11c. An algebraic approximation to this curve, valid within 1 db over the range $2 \leq B_R/B_B \leq 200$ (or $0 \leq D \leq 99$), is given by

$$K_{Rt} = \sqrt{5\ B_R/B_B} \cong \sqrt{10(D + 1)} \tag{199a}$$

where the approximation in terms of D follows from Eq. (198). The corresponding value of received RF signal is

$$S_{Rt} = K_{Rt}\ B_R\ N_o = 2\sqrt{10}(D+1)^{3/2}\ B_B\ N_o \tag{199b}$$

For a given deviation ratio D, this is the minimum value of S_R for which the FM expressions from Table 5 such as

$$S_R = \frac{\Lambda_B}{3D^2}\ B_B\ N_o\ K_L \tag{200}$$

are valid.

*This is by no means a generally accepted criterion. Constant values of 10 (10 db) and 16 (12 db) are also commonly cited for K_{Rt}.

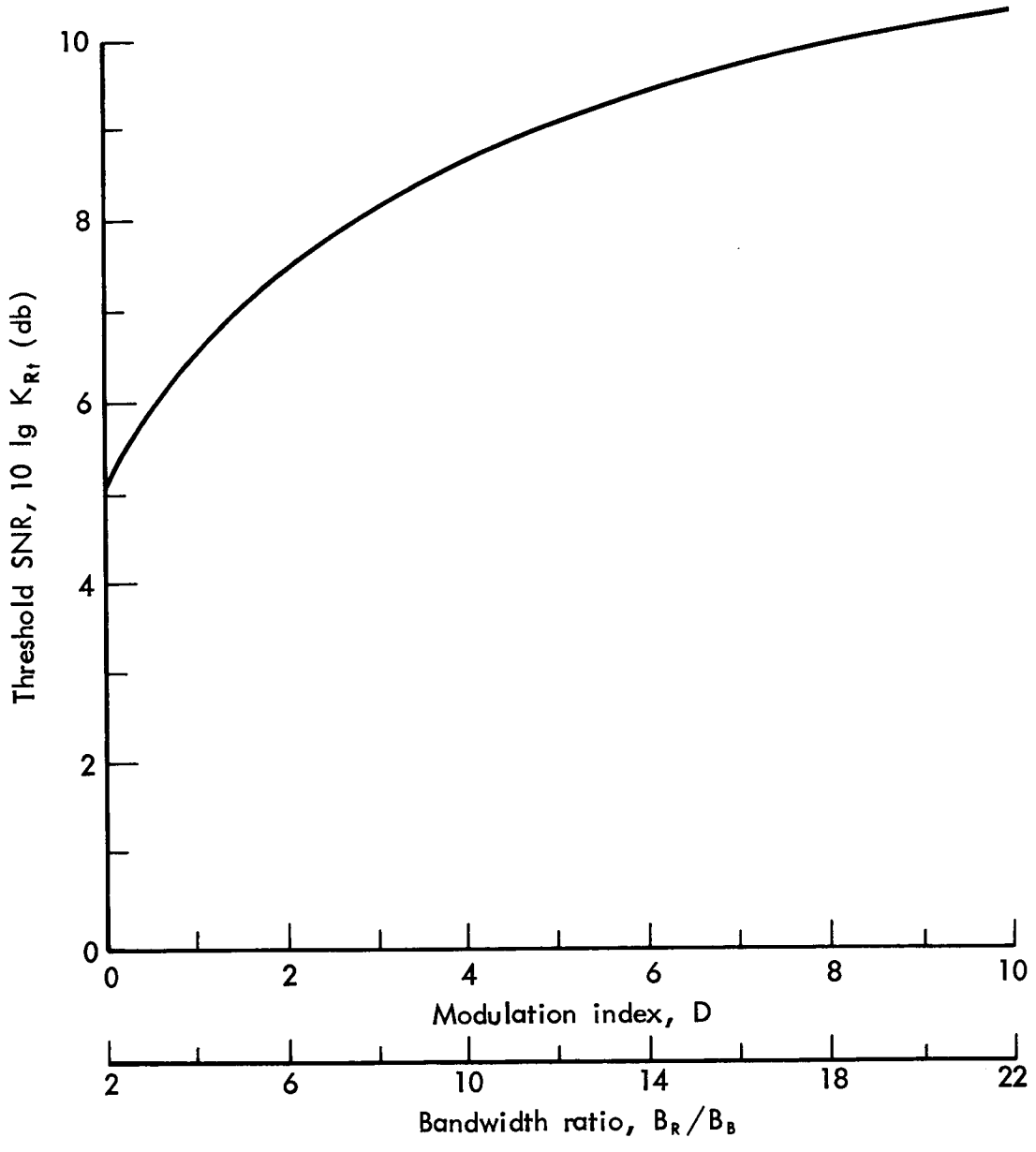


Fig. 11c—Threshold SNR for frequency demodulator without feedback

To operate with this minimum received signal, the FM modulation index D is merely chosen to make S_R equal to S_{Rt}. Denoting this value of modulation index by D_t, it follows from Eqs. (199b) and (200) that

$$D_t^2 (D_t^2 + 1)^{3/2} = \frac{\Lambda_B K_L}{6\sqrt{10}} = 0.0527 \Lambda_B K_L \tag{201a}$$

Thus, the deviation ratio for threshold operation depends only on the desired output signal quality specified by K_L and on the peak-to-average power ratio, Λ_B, of the baseband signal. In most cases the product of these quantities will be so large that D_t will be much greater than unity, and Eq. (201a) can be replaced by

$$D_t \cong \left(\frac{\Lambda_B K_L}{6\sqrt{10}}\right)^{2/7} = 0.431 (\Lambda_B K_L)^{2/7} \qquad \Lambda_B K_L >> 1 \tag{201b}$$

Substituting this value into Eq. (200) yields an expression for the threshold value of received carrier power directly in terms of the SNR objective K_L, the receiver noise density N_o, and the baseband parameters B_B and Λ_B:

$$S_{Rt} \cong 5.38 B_B N_o (\Lambda_B K_L)^{3/7} \qquad \Lambda_B K_L >> 1 \tag{201c}$$

The threshold condition sets a limit on the extent to which carrier power can be reduced in exchange for increased RF bandwidth using a conventional FM demodulator. When operating at threshold, carrier power is at a minimum S_{Rt}, and RF bandwidth B_R is at a maximum. In practice, there are at least two reasons why threshold operation might not be used.

The first occurs when the received signal power fluctuates or fades due to varying path length, to varying absorption in the medium, or to other causes. To avoid serious deterioration of the output signal during such fades it is necessary to use a transmitter power high enough so that the received carrier is above threshold most of the time (99.9 percent of the time, for example). A second obvious reason for not operating at threshold is that the corresponding RF bandwidth may exceed that allocated for the circuit in question. If this is the case, the maximum deviation ratio is by Eq. (198)

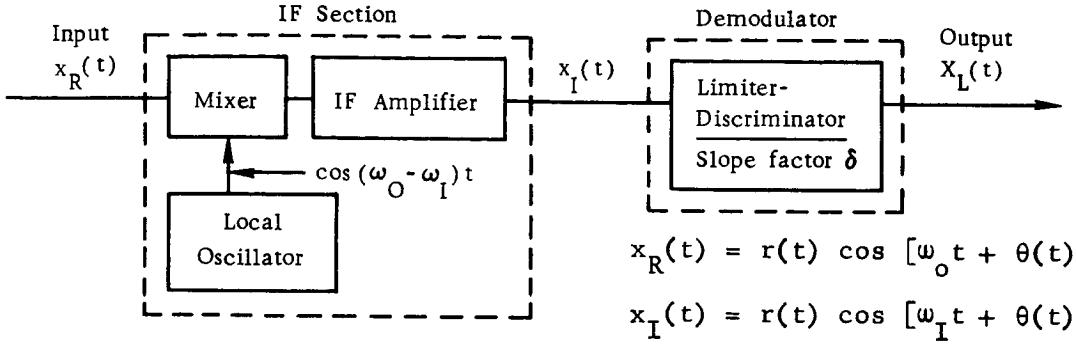
$$D_a \cong \frac{B_{Ra}}{2B_B} - 1 \tag{201d}$$

where B_{Ra} is the RF bandwidth allocated to the link in question.

Feedback FM Receiver

The FM carrier-to-noise power ratio at threshold, and with it the minimum carrier power required for a given output SNR, can be reduced below the values just discussed for a conventional FM receiver by using the technique of frequency feedback first described by Chaffee.[20] As a prelude to describing this technique, it will be useful to recapitulate the operation of a conventional superheterodyne FM receiver.

Such a receiver consists of an IF section and a frequency demodulator (limiter-discriminator) as shown in the block diagram of Fig. 12a. It will be recalled that the IF section plays no direct role in signal recovery. The frequency of the local oscillator is fixed and the mixer serves only to translate the spectrum of the input signal-plus-noise waveform from the vicinity of the carrier angular frequency ω_o to the

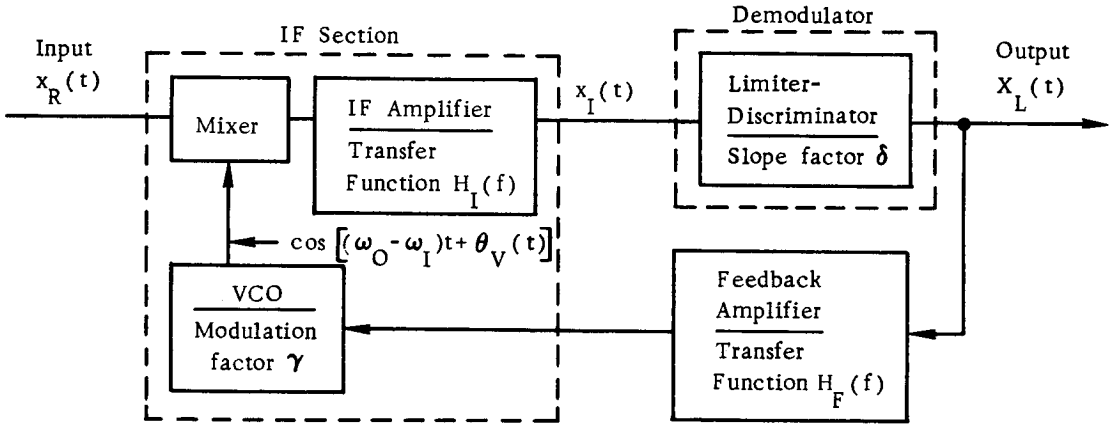


a. Conventional FM receiver

$$x_R(t) = r(t) \cos [\omega_o t + \theta(t)]$$

$$x_I(t) = r(t) \cos [\omega_I t + \theta(t)]$$

$$X_L(t) = \delta \dot{\theta}(t) \cong \delta[\dot{\varphi}(t) + \dot{q}(t)]$$

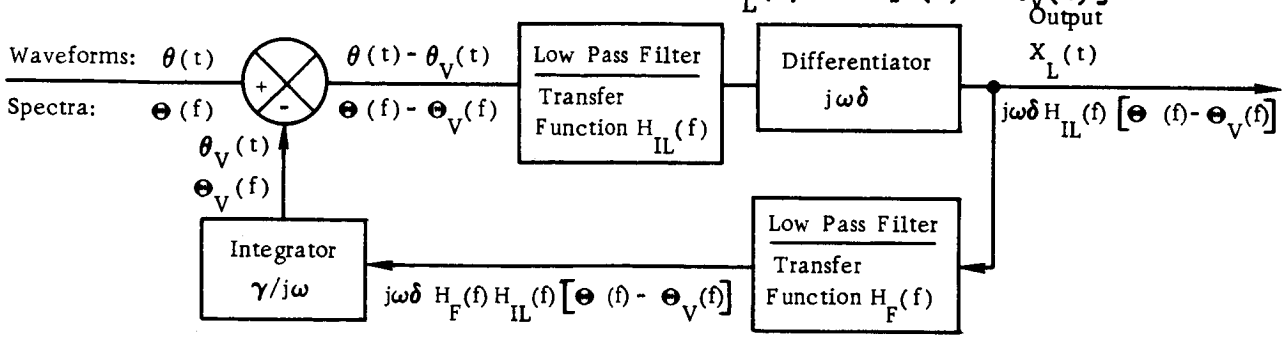


b. Feedback FM receiver

$$x_R(t) = r(t) \cos [\omega_o t + \theta(t)]$$

$$x_I(t) \cong r(t) \cos [\omega_I t + \theta(t) - \theta_V(t)]$$

$$X_L(t) \cong \delta [\dot{\theta}(t) - \dot{\theta}_V(t)]$$



c. Baseband analog of feedback FM receiver

Fig. 12—FM receiver block diagrams

vicinity of the IF angular frequency ω_I. The essentially rectangular band-pass filters of the IF amplifier have a bandwidth B_I equal to the bandwidth B_R of the modulated RF carrier, and serve only to select from the mixer output the spectral band about ω_I which contains the translated spectrum of the modulated carrier. One consequence of this is that the normalized signal-plus-noise waveform at the output of the IF section

$$x_I(t) = e_I(t) + u_I(t) = r(t) \cos [\omega_I t + \theta(t)]$$

differs from its counterpart $x_R(t)$ at the receiver input only in the replacement of ω_o by ω_I. Another consequence is that the SNR, K_I, at the demodulator input is the same as K_R, the SNR at the receiver input. From the waveform $x_I(t)$ at its input, the frequency demodulator then extracts an output $x_L(t)$ which is proportional to the instantaneous frequency deviation $\dot{\theta}(t)$. As long as the input carrier-to-noise ratio K_R exceeds the threshold value K_{Rt}, Eqs. (189) and (190) show that $\dot{\theta}(t)$ is given by

$$\dot{\theta}(t) \cong \dot{\varphi}(t) + \dot{q}(t) \tag{202}$$

where $\dot{\varphi}(t)$ is the angular frequency deviation imposed at the transmitter and $q(t)$ is the quadrature component of the normalized input noise waveform introduced in Eq. (129). The output SNR, $K_L = \overline{\dot{\varphi}^2(t)}/\overline{\dot{q}^2(t)}$, is then given in terms of the input power, the receiver noise density, and the parameters of the baseband and modulator by Eq. (197b).

In comparison with a conventional receiver, the feedback FM receiver* depicted in Fig. 12b includes a feedback amplifier which uses

*The terms "frequency feedback demodulator" or "frequency compression demodulator" are also used.

the discriminator output to frequency modulate the local oscillator, which is now a VCO (voltage controlled oscillator).* The polarity of the control voltage from the feedback amplifier is arranged so as to reduce the frequency deviation of the IF signal, i.e., the feedback is negative. In particular, the peak frequency deviation and hence the deviation ratio and bandwidth of the IF waveform can be made much smaller than those of the modulated RF carrier. This permits the IF amplifier bandwidth to be made significantly narrower than the RF signal bandwidth B_R, although in order for the feedback to be stable the IF band pass can no longer be made rectangular.

A quantitative description of the operation of a feedback FM receiver is probably most easily given in terms of the baseband analog circuit shown in Fig. 12c. Here, the mixer is replaced by a subtractor, the IF amplifier by its low-pass equivalent filter, the discriminator by an ideal differentiator, and the VCO by an ideal integrator. The utility of this representation stems from the fact that when the peak phase deviation at IF is small (as when the signal is narrow-band, or when the phase rate is small and the signal is quasi-stationary) the output $X_L(t)$ of the actual receiver is closely approximated by the response of the baseband analog circuit to the instantaneous phase deviation $\theta(t)$ of the RF input signal-plus-noise waveform $x_R(t)$.

The feedback FM receiver is characterized by the closed-loop transfer function

---

*A VCO is an oscillator whose frequency is proportional to the control voltage.

$$H_c(f) \equiv \Theta_v(f)/\Theta(f) \tag{203}$$

which relates the spectrum of the VCO output phase deviation $\theta_v(t)$ to that of the phase deviation $\theta(t)$ of the RF input $x_R(t)$. To express $H_c(f)$ in terms of the receiver parameters, the mixer output spectrum $\Theta(f) - \Theta_v(f)$, in Fig. 12c, is traced around the feedback loop to the integrator output yielding the relation

$$\Theta_v(f) = \gamma\delta\ H_F(f)\ H_{IL}(f)\ [\Theta(f) - \Theta_v(f)]$$

where

γ = modulation factor of VCO

δ = deviation sensitivity of discriminator

$H_F(f)$ = transfer function of feedback amplifier

$H_{IL}(f)$ = transfer function of low-pass equivalent of IF amplifier

When combined with the definition of $H_c(f)$, the result is

$$H_c(f) = \frac{\gamma\delta\ H_F(f)\ H_{IL}(f)}{1 + \gamma\delta\ H_F(f)\ H_{IL}(f)} \tag{204}$$

The closed-loop transfer function is seen to be closely related to the open-loop transfer function

$$H_o(f) = \gamma\delta\ H_F(f)\ H_{IL}(f) \tag{205}$$

defined by the ratio $\Theta_v(f)/\Theta(f)$ when the feedback loop is opened at the oscillator input to the mixer.

Both of these transfer functions and the noise bandwidths B_{Nc} and B_{No} associated with them* are important in describing the properties of a feedback FM receiver. From an operational standpoint, however, the most important parameter is the feedback factor F defined as the ratio of the instantaneous frequency deviation at RF to that at IF. Since the frequency deviation is proportional to the derivative of the phase deviation, the feedback factor may also be defined as

$$F = \frac{\theta(t)}{\theta(t) - \theta_v(t)} \equiv \frac{1}{1 - \theta_v(t)/\theta(t)} \tag{206}$$

To interpret F in terms of circuit parameters, observe that when $\theta(t)$ is sinusoidal with frequency f, Eqs. (203) to (205) combine with Eq. (206) to give

$$F = \frac{1}{1 - H_c(f)} = 1 + H_o(f)$$

$$= 1 + \gamma\delta\, H_F(f)\, H_{IL}(f)$$

In a well-designed feedback FM receiver, the transfer function product $H_F(f)\, H_{IL}(f)$ is chosen so as to make the feedback factor, F, virtually constant over the entire range of baseband frequencies. In numerical calculations, F will usually be expressed in db

$$F_{db} = 20 \log F = 20 \log \left|1 + \gamma\delta\, H_F(f)\, H_{IL}(f)\right|$$

---

*The noise bandwidth B_N of a symmetrical filter with transfer function H(f) is the bandwidth of the rectangular filter with amplitude response $|H(0)|$ that passes the same white noise power. In general, $|H(0)|^2 B_N = \int_{-\infty}^{\infty} |H(f)|^2 dt$.

because from a feedback amplifier point of view it represents the signal gain of the feedback loop.

In considering the use of feedback reception on an FM link, it is quite important to recognize that the performance of a feedback FM receiver above threshold is described by exactly the same relation, Eq. (197b), that applies to a conventional FM receiver. Put another way, for a given baseband and receiver sensitivity, the output SNR, K_L, depends only on the deviation ratio, D, and received power, S_R, of the modulated carrier and not on the feedback factor, F, of the receiver. The reason may be seen from Eqs. (202) and (206) which show that feedback reduces the frequency deviation of both the signal and noise components at the discriminator input by the same factor. Thus while both the signal and noise power at the discriminator output are reduced, the SNR remains unchanged from the value given by Eq. (197b).

Another quantity that is not changed by moderate amounts of feedback is the threshold value K_{It} of carrier-to-noise ratio measured at the input to the frequency demodulator. As with the frequency demodulator of a conventional receiver, threshold effects begin to appear when the IF signal-to-noise ratio approaches

$$K_{It} \cong \sqrt{5\ B_I/B_B} \tag{207}$$

where B_I is the noise bandwidth of the IF amplifier.

With the conventional receiver, however, the SNR and bandwidth at IF are equal to those at RF, and this threshold condition yields the previously stated result

$$K_{Rt} \equiv \frac{S_{Rt}}{N_o B_R} \cong \sqrt{5\ B_R/B_B} \cong \sqrt{10\ (D + 1)} \qquad (199a)$$

in which Carson's rule

$$B_R \cong 2\ B_B\ (D + 1) \qquad (198)$$

is used to express the RF bandwidth in terms of the modulation index D of the modulated carrier.

In the case of a feedback FM receiver with feedback factor, F, the deviation ratio at IF is reduced to D/F and the noise bandwidth of the IF amplifier can be reduced to the value given by Carson's rule[21]

$$B_I \cong 2\ B_B \left(\frac{D}{F} + 1\right)$$

Substituting this into Eq. (207), and assuming that the input noise density N_o is unchanged by feedback, the threshold carrier-to-noise ratio at the input of a feedback FM receiver becomes

$$K_{Rt} \equiv \frac{S_{Rt}}{N_o B_R} = \frac{B_I}{B_R}\ K_{It} = \sqrt{10}\,\frac{\left(\frac{D}{F} + 1\right)^{3/2}}{D + 1} \qquad (208a)$$

Note that in the absence of feedback (F = 1) this equation reduces to Eq. (199a) as it should. The RF signal power at threshold is also of interest and may be obtained by combining Eqs. (198) and (208a)

$$S_{Rt} = 2\sqrt{10}\left(\frac{D}{F} + 1\right)^{3/2} \qquad (208b)$$

It will be observed that this reduces to Eq. (199b) when F = 1.

Equations (208a) and (208b) respectively give the values of K_{Rt} and S_{Rt} corresponding to threshold operation with a specified modulation index D and feedback factor F. To find the modulation index, D_t, corresponding to threshold operation at a given output SNR, K_L, either equation may be combined with Eq. (197b) to yield the feedback counterpart of Eq. (201a):

$$D_t^2 \left(\frac{D_t}{F} + 1\right)^{3/2} = \frac{\Lambda_B K_L}{6\sqrt{10}} \tag{209}$$

Note that for a given baseband peak-to-average ratio, Λ_B, and output SNR, K_L, the threshold modulation index, D_t, increases with increasing feedback factor, F. In this way feedback extends the range over which RF bandwidth can be traded for RF power in accordance with Eq. (197b).

Unfortunately there is a limit to the amount of feedback that can be used. As determined experimentally and explained theoretically by Enloe,[22] there is a second threshold, in addition to that given by Eq. (207), which occurs when the rms phase deviation of the VCO output caused by noise increases to about one-third of a radian. Expressed in terms of the carrier-to-noise ratio in the closed-loop noise bandwidth B_{Nc}, this "feedback threshold" condition becomes[22]

$$\frac{S_{It}}{N_o B_{Nc}} = 4.8 \left(\frac{F - 1}{F}\right)^2 \tag{210}$$

In the development of this expression it is assumed that the conventional threshold condition of Eq. (207) is already met. This simply means that the IF carrier power S_I must exceed the larger of the two values S_{It} given by Eqs. (207) and (210). The maximum feedback factor, F_m, is that for which these two threshold values are equal.

In order to determine this maximum feedback factor it is first necessary to express the closed-loop noise bandwidth in terms of the baseband bandwidth and the feedback factor. Plots of B_{Nc}/B_B versus phase margin are given in Fig. 8 of Ref. 22 for various values of F, assuming a Bode-type open=loop transfer function obtained with single-pole filters in both the feedback and IF amplifiers. For F<30 db, the values corresponding to a phase margin of about 50^o are well represented by the equation

$$B_{Nc} = 3.1\, B_B\, F^{0.59}$$

Substituting this expression for B_{Nc} into Eq. (210), and solving the result simultaneously with Eq. (208a) to eliminate S_{It}, yields the desired expression for F_m

$$F_m \left[1.77 \left(\frac{F_m - 1}{F_m}\right)^{4/3} F_m^{.39} - 1\right] = D \qquad (211)$$

A plot of F_m versus D based on this equation is given in Fig. 13.

Using the values of F_m given by Fig. 13 in conjunction with the feedback threshold conditions previously derived, the complete power-bandwidth tradeoffs for FM can be summarized. This is done in Fig. 14, which is basically a plot of Eq. (197b). The solid straight lines show the combinations of received carrier power S_R and RF bandwidth B_R required for an output SNR, K_L, when FM is used to transmit a baseband of bandwidth B_B and peak-to-average power ratio Λ_B to a receiver whose input noise power density is N_o. An auxiliary scale shows the modulation index D required in each case. The dashed lines then show the combinations of power S_{Rt} and modulation index D_t corresponding to threshold operation with various amounts of feedback F as given

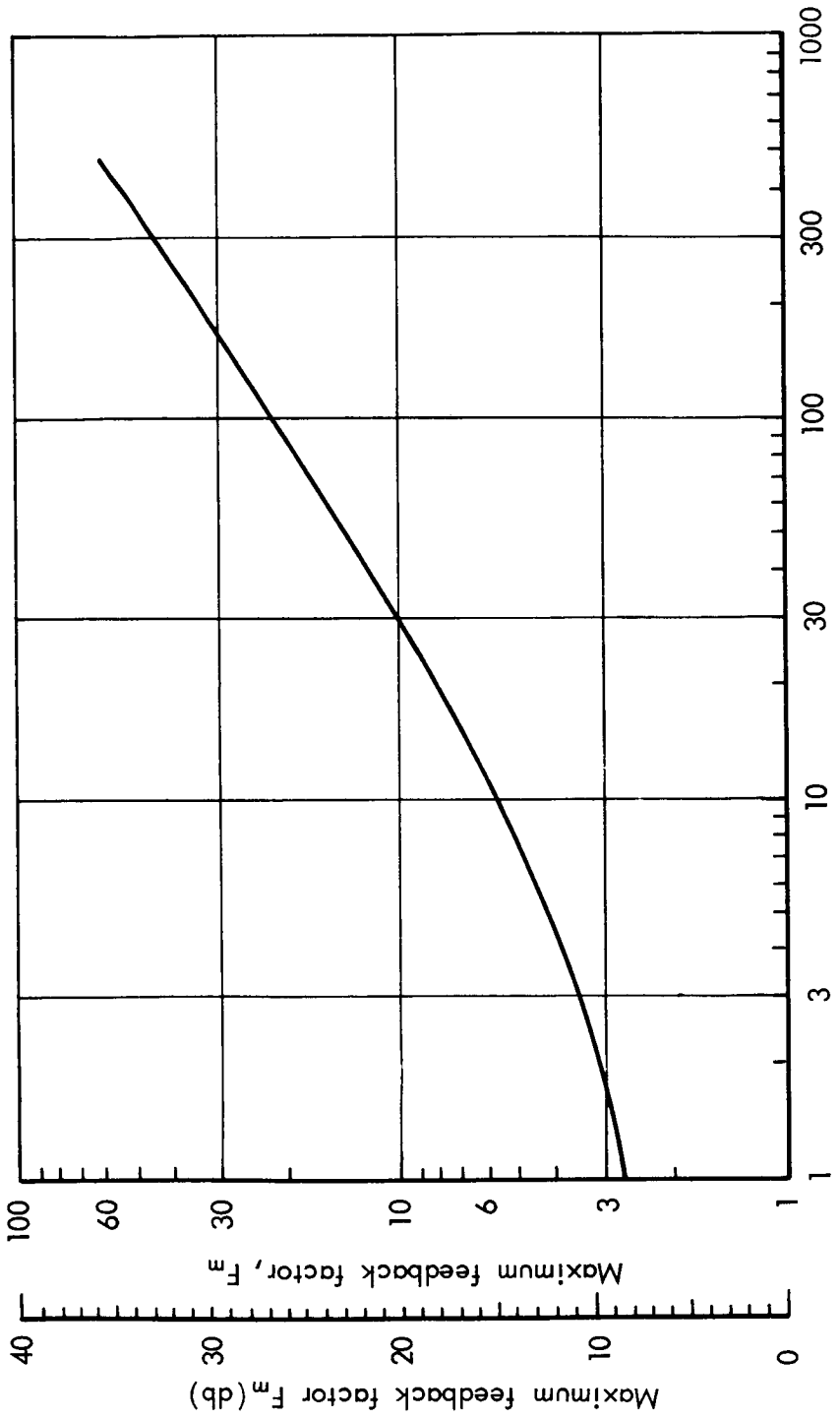


Fig. 13—Feedback factor for which conventional and feedback threshold conditions are satisfied simultaneously

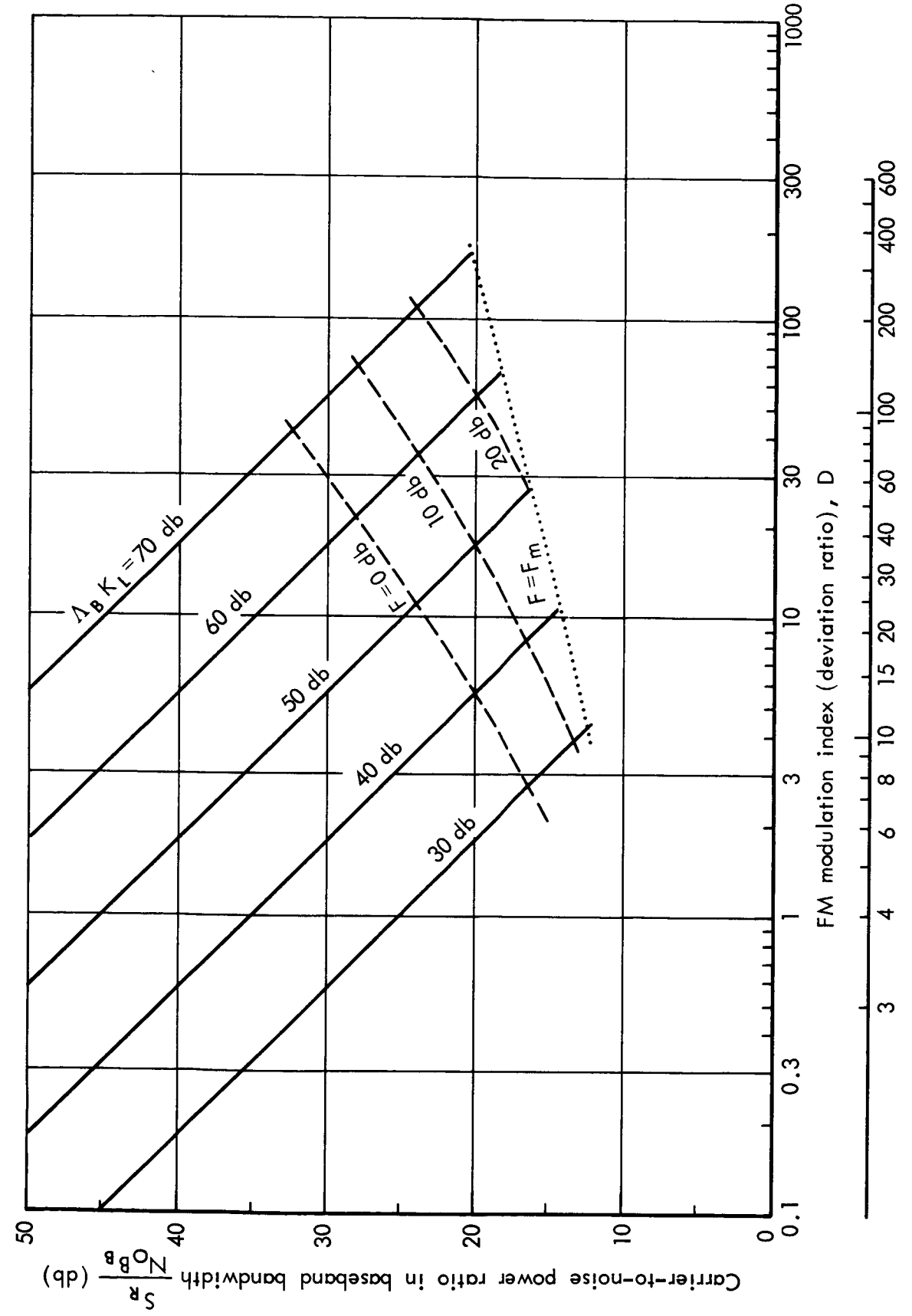


Fig. 14—FM power-bandwidth tradeoffs

by Eqs. (208b) and (209). In particular, the curve labeled F = 0 db shows the threshold without feedback. Finally, the dotted line which terminates the solid lines corresponds to threshold operation with the maximum amount of feedback as described by Eq. (211) and Fig. 13.

To illustrate these relationships, suppose that a baseband which occupies a bandwidth of 15 kc and has a peak-to-average power ratio of 10 is to be recovered with an output SNR of 40 db over an FM link with an input noise power density of 4×10^{-18} W/kc.* Referring to Fig. 14, the combinations of received carrier power and RF bandwidth that may be used are indicated by the solid line labeled $\Lambda_B B_B = 50$ db.

For example, with a modulation index of 3, the required carrier-to-noise ratio in the baseband bandwidth is 35.7 db and the ratio of RF to baseband bandwidth is 8. Thus, the RF bandwidth is 120 kc, and, since $N_o B_B = 6 \times 10^{-17}$ W (-162.2 dbw), the carrier power at the receiver input must be -162.2 + 35.7 = -126.5 dbw.

Alternatively, if the RF spectrum is available, the FM modulation index may be increased to any value up to 11 (which corresponds to the threshold of a conventional FM receiver (F = 0 db)). With this index the RF bandwidth is 24 x 15 = 360 kc and the input carrier power is -162.2 + 24.3 = -137.9 dbw. Compared with operation at a modulation index of 3, RF power has been reduced by 11.4 db at the cost of a three-fold increase in bandwidth.

Finally, if a feedback FM receiver is used, it is seen from Fig. 14 that feedback up to a maximum slightly in excess of 20 db may be used.

---

*This corresponds to a system noise temperature of 290°K--see Eq. (134).

With this amount of feedback, the RF bandwidth is 58 x 15 = 870 kc but the input carrier power is reduced to only -162.2 + 16.3 = -145.9 dbw, representing an 8 db reduction over threshold operation without feedback. For comparison, operating the link with a SSB transmitter and receiver would require an RF bandwidth of only 15 kc but the average carrier power would be -162.2 + 40 = -122.2 dbw, as may be seen with the aid of the SSB relations in Tables 3 and 5. Moreover, the SSB transmitter would have to have a peak power capability 10 to 12 db higher than this, depending on the baseband statistics.

III. FREQUENCY-DIVISION MULTIPLEXING

INTRODUCTION

As previously noted, the term multiplexing refers to the process of combining a number of independent signals to form a single composite signal called a baseband. In baseband form, the multiple signals can be sent by a single transmitter to a single receiver, at whose output the individual signals are recovered by subjecting the baseband to the inverse process of demultiplexing. The locations of the multiplexer and demultiplexer in the communications link were shown in Fig. 1.

In this section the principal analog-multiplexing technique will be discussed with particular attention to the amplitude characteristics of the basebands that result when input signals are not processed prior to multiplexing. The effects of signal processing on the baseband will then be discussed in the next section.

Multiplexing techniques can be categorized according to whether the individual signals are separated in frequency or in time.* In frequency-division multiplexing (FDM), each channel is assigned a separate band of frequencies, and the baseband spectrum consists of the totality of the single channel bands. Each input signal is placed in its assigned frequency "slot" in the baseband spectrum by using it to modulate a sinusoid of appropriate frequency called a subcarrier (to distinguish it from the single RF carrier which the baseband itself modulates). Demultiplexing then consists of filtering the single channel slots from the baseband spectrum at the receiver output and individually demodulating the subcarriers.

---

*Time-division multiplexing is covered in a companion Memorandum.(2)

The characteristics of the FDM baseband signal, in both amplitude and frequency, depend of course not only on the type of signals being multiplexed but on the type of subcarrier modulation used. With analog input signals, single-sideband suppressed carrier modulation (SSB) is used almost universally. Its principal advantages are that (1) when a given channel is not in use, it makes no contribution to the baseband, and (2) the bandwidth B_C of the channel need be little greater than the bandwidth B_S of the single channel signal input that it carries. The bandwidth B_B of the FDM baseband is equal to the total number Q of channels it comprises times the channel bandwidth; i.e.,

$$B_B = QB_C$$

In the case of voice signals, for example, the standard values are $B_S = 3.1$ kc, $B_C = 4$ kc.

Frequency modulation (FM) has also been used to place single channel analog signals on FDM subcarriers, but the greater bandwidths required and the care which must be exercised in selecting subcarrier frequencies to minimize crosstalk between subcarriers* have inhibited its popularity for telephone use.

With narrow-band digital input signals such as teletype, a digital-modulation technique such as frequency shift keying or on-off keying** of appropriately spaced audio-frequency subcarriers is normally used to place the input signals in their proper frequency slots in an FDM baseband having the bandwidth of a single voice channel. In this form, the multiplexed teletype signals may be treated as any other voice channel; in particular it may be frequency-division multiplexed with other voice channels using SSB subcarrier modulation.

---

*Note that with FM, all of the subcarriers are present even at times when some channels are carrying no signal.

**See Ref. 2 for a discussion of these techniques.

SIGNAL-TO-NOISE RATIOS

When the modulating signal or baseband is a multiplex of two or more channels, the signal-to-noise ratio K_L at the receiver output is no longer of direct interest. What is of interest is the quality of the individual signals at the single channel outputs of the demultiplexer. This quality is measured by the single channel output signal-to-noise ratios K_{Ci}, where the subscript $i = 1,2,\ldots, Q$ distinguishes the different channels which make up a Q channel baseband. Of course, the K_{Ci} and K_L are closely related; and it is through K_L that K_{Ci} can be expressed in terms of K_R, the RF signal-to-noise ratio at the receiver input.

Such expressions are easy to obtain for FDM basebands with the aid of the signal-to-noise expressions developed for different modulation methods in the preceding section and tabulated in Table 5. Assuming that the subcarriers are SSB modulated, the demultiplexer will consist of Q product demodulators, each furnished with a locally generated subcarrier having the proper frequency and phase. If f_i is the subcarrier frequency for the i^{th} channel, and if upper sidebands are used, the input $X_{Li}(t)$ to the i^{th} channel demodulator is obtained by filtering the band $f_i < f < f_i + B_{Ci}$ from the receiver output $X_L(t)$. Since the input and output signal-to-noise ratios are equal for product demodulation of SSB, it follows that

$$K_{Ci} = K_{Li} \tag{212}$$

where

$$K_{Li} \equiv S_{Li}/N_{Li} \equiv \frac{S_{Li}/S_L}{N_{Li}/N_L} K_L \tag{213}$$

is the ratio of signal-to-noise power in the i^{th} demodulator input $X_{Li}(t)$,

and $K_L \equiv S_L/N_L$ is the corresponding ratio for the receiver output $X_L(t)$. Thus, to express K_{Ci} in terms of K_L it is only necessary to evaluate the ratios S_{Li}/S_L and N_{Li}/N_L which appear in Eq. (213).

The first of these ratios does not depend on the type of modulation used to transmit the baseband. The receiver output signal $E_L(t)$ is proportional to the baseband signal $E_B(t)$ for all modulation methods and hence the ratio of the signal power in the i^{th} channel to the total signal power is the same at the receiver output as at the transmitter input--i.e.,

$$S_{Li}/S_L = S_{Bi}/S_B \equiv 1/\Gamma_i \tag{214}$$

The ratio Γ_i of the average power in the baseband to that contained in the i^{th} channel is called the baseband load factor (relative to the i^{th} channel).

The noise ratio N_{Li}/N_L does depend on modulation method, however, because $W_{U_L}(f)$, the noise power spectrum at the receiver output, can be frequency dependent. In general,

$$N_{Li}/N_L = \int_{f_i}^{f_i + B_C} W_{U_L}(f)\,df \Big/ \int_0^{B_B} W_{U_L}(f)\,df \tag{215}$$

Expressions for $W_{U_L}(f)$, derived for different demodulators in the preceding section, and the resultant values of N_{Li}/N_L and K_{Ci}/K_L, obtained with the aid of Eqs. (212) and (215), are summarized in Table 6.

The last-named ratios may then be combined with those in Table 5 to yield expressions which give directly the RF signal-to-noise ratio K_R, and the RF signal powers S_R and $\hat{S}_R$ required to produce a specified

Table 6

BASEBAND OUTPUT NOISE POWER SPECTRA, SINGLE CHANNEL OUTPUT NOISE POWER, AND SINGLE CHANNEL OUTPUT SNR

Modulation Method	Output Noise Power Spectrum $W_{U_L}(f)$	Single Channel Noise Power N_{Li}/N_L	Single Channel Output SNR K_{Ci}/K_L
AM	$\frac{1}{4} G_I N_o$	$\frac{B_{Ci}}{B_B}$	$\frac{B_B}{\Gamma_i B_{Ci}}$
DSB			
SSB			
PM	$\frac{1}{8} N_o/S_{Ro}$		
FM	$\frac{1}{8} \omega^2 N_o/S_{Ro}$	$\sim 3\left(\frac{f_i}{B_B}\right)^2 \frac{B_{Ci}}{B_B}$	$\sim \frac{1}{3}\left(\frac{B_B}{f_i}\right)^2 \frac{B_B}{\Gamma_i B_{Ci}}$
SSBFM	$\frac{1}{12} \omega^2 N_o/S_{Ro}$		

$\Gamma_i \equiv S_B/S_{Bi}$ = baseband load factor

value of single channel SNR K_{Ci}. The results of this operation are displayed for each of the analog modulation methods in Table 7. Note that the entries for FM and SSBFM in both Tables 6 and 7 are approximations for the commonly occurring special case where the bandwidth B_{Ci} of the i^{th} channel is small compared with the i^{th} subcarrier frequency f_i. The general result may be obtained by replacing $\frac{\frac{1}{3} B_B^3}{f_i^2 B_{Ci}}$ by $\frac{B_B^3}{(f_i+B_{Ci})^3 - f_i^3}$.

SPEECH BASEBANDS

One of the principal reasons for the use of multiplexing on radio circuits has already been mentioned--that is, it permits the use of only a single RF transmitter and receiver. This is an advantage despite the fact that at terminals (as opposed to relay stations) almost as much equipment is required to multiplex and demultiplex the baseband as would be required for independent RF transmission and reception of the individual channels. Thus, for each channel in an FDM baseband, there is a small transmitter (modulator) in the multiplexer and a small receiver (demodulator) in the demultiplexer. The advantages of multiplexing stem from the fact that it can be done at comparatively low frequencies and powers and, at relay stations, there is no need for multiplexing equipment at all unless channels are to be removed from or added to the baseband.

Another basic reason for the use of multiplexing techniques is to improve the efficiency of transmission for human speech. This improvement is made possible by: (1) the unique amplitude distribution of speech, (2) the fact that not all channels on a multichannel telephone circuit are busy simultaneously, and (3) the fact that even when a telephone channel is busy, it does not carry speech continuously. The result

Table 7

INPUT SNR AND INPUT SIGNAL POWER VERSUS SINGLE CHANNEL OUTPUT SNR

Modulation Method	RF SNR K_R/K_{Ci}	Average RF Power $\frac{S_R/B_{Ci}N_o}{K_{Ci}}$	Peak RF Power $\frac{\hat{S}_R/B_{Ci}N_o}{K_{Ci}}$
AM	$\left(1 + \frac{\Lambda_B}{M^2}\right) \frac{1}{2} \Gamma_i B_{Ci}/B_B$	$\left(1 + \frac{\Lambda_B}{M^2}\right) \Gamma_i$	$\left(1 + \frac{1}{M}\right)^2 2\Lambda_B \Gamma_i$
DSB	$\frac{1}{2}\Gamma_i B_{Ci}/B_B$	Γ_i	$2\Lambda_B \Gamma_i$
SSB[a]	$\Gamma_i B_{Ci}/B_B$	Γ_i	$k_1 \Lambda_B \Gamma_i$
PM	$\frac{\Lambda_B}{\Phi^2} \Gamma_i B_{Ci}/B_R$	$\frac{\Lambda_B}{\Phi^2} \Gamma_i$	$2 \frac{\Lambda_B}{\Phi^2} \Gamma_i$
FM	$\sim \frac{\Lambda_B}{D^2} \left(\frac{f_i}{B_B}\right)^2 \Gamma_i B_{Ci}/B_R$	$\sim \frac{\Lambda_B}{D^2} \left(\frac{f_i}{B_B}\right)^2 \Gamma_i$	$\sim 2 \frac{\Lambda_B}{D^2} \left(\frac{f_i}{B_B}\right)^2 \Gamma_i$
SSBFM[a]	$\sim \frac{\Lambda_B}{D^2} \left(\frac{f_i}{B_B}\right)^2 \Gamma_i B_{Ci}/B_R$	$\sim 2/3 \left[\frac{\Lambda_B}{D^2} \left(\frac{f_i}{B_B}\right)^2 \Gamma_i\right]$	$\sim 2/3\ k_2 \left[2 \frac{\Lambda_B}{D^2} \left(\frac{f_i}{B_B}\right)^2 \Gamma_i\right]$

$\Gamma_i \equiv \overline{S}_B/\overline{S}_{Bi}$ = baseband load factor

[a]The factors k_1 and k_2 represent the ratio Λ_R/Λ_B for SSB and SSBFM respectively. (See Table 3)

is that when Q speech channels are combined into an FDM baseband using SSB subcarrier modulation, the transmitter power required is considerably less than Q times the transmitter power that would be required to transmit one of the channels using the same RF modulation method and modulation index. With the angle modulation methods, it is also possible to trade some of this power saving for a reduction in the RF bandwidth required to transmit the Q speech channels.

The balance of this section is devoted to an examination of this type of frequency-division multiplexed speech baseband, with the object of determining the magnitude of the power and/or bandwidth savings that multiplexing makes possible.*

Single Channel Inputs

The dynamic characteristics of any baseband formed by multiplexing a number of speech channels are obviously dependent on the characteristics of the component single speech channels. For this purpose such channels can be described by two probability distributions.

The first (Fig. 15) shows how the instantaneous speech power $E_{Si}^2(t)$ in a particular channel is distributed about its average value S_{Si} at a time when the channel is "active", i.e., during an interval when the channel is carrying ordinary connected speech, including the short pauses between syllables and words. During a telephone conversation, two channels--one in each direction--are "busy" (unavailable for use by other subscribers), but neither channel is active more than a fraction of

*The approach followed here is similar to that of Holbrook and Dixon.[23] An alternative approach using Monte Carlo methods was recently put forth by Cyr and Thuswaldner.[24]

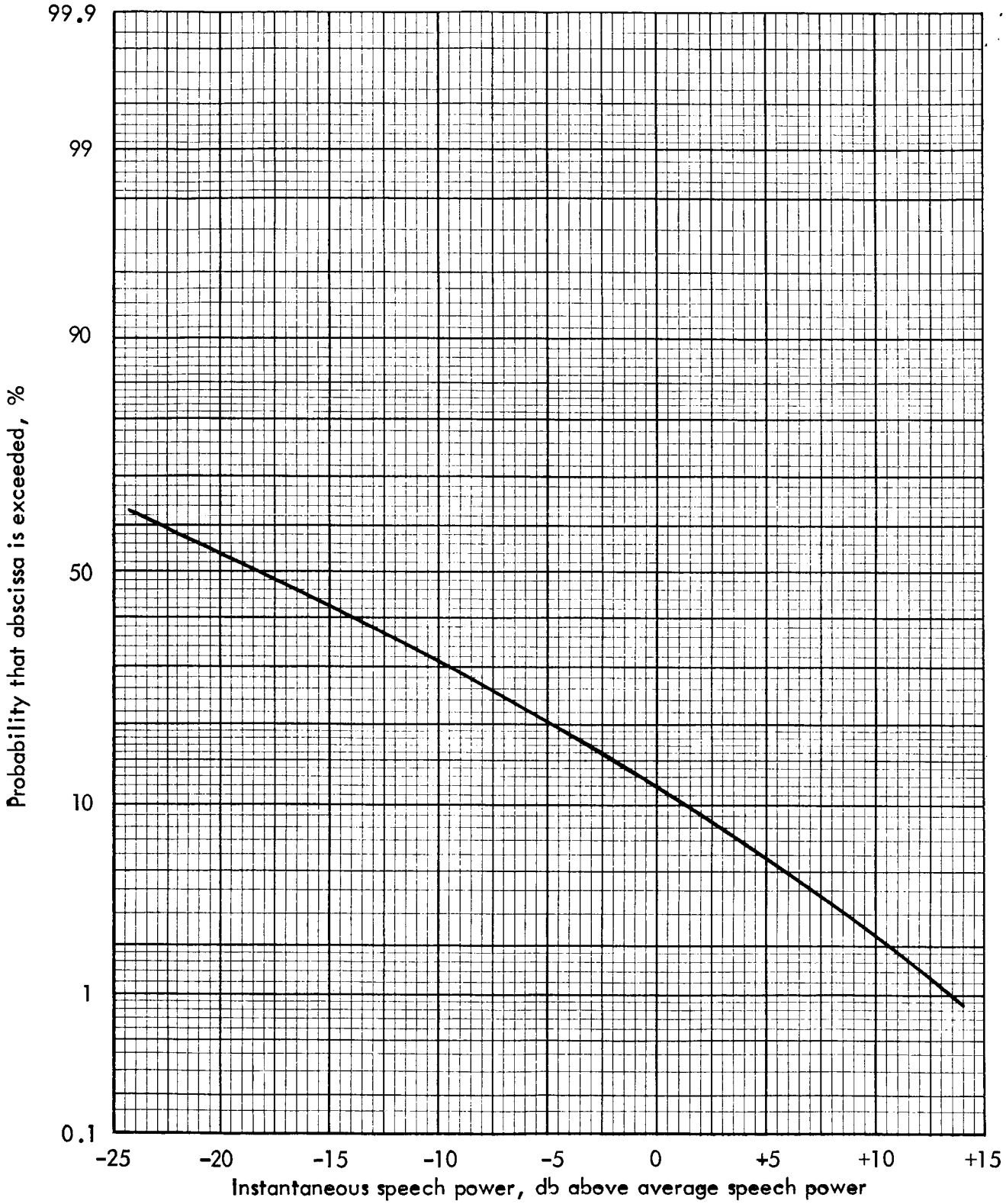


Fig. 15—Distribution of instantaneous speech power for a single speaker

this time. In particular, a busy channel is not active for a good portion of the period when a telephone connection is being established, nor during pauses in the conversation and when one party is listening to the other. This is important because with SSB modulation used to form the FDM baseband, a channel contributes to the baseband signal only when it is active.

Experimental measurements reported by Holbrook and Dixon[23] and others[25-27] show that the distribution given in Fig. 15 adequately represents the variations of instantaneous speech power in an active channel for the majority of talkers and for most types of telephone instruments. Equally important, the same distribution relative to the average speech power applies regardless of the value of this average, so long as it is not so high that the telephone instrument overloads on the peak speech voltages.

Of course, the average speech power carried by an active channel will itself vary from one active channel to another at a particular time, and from time to time on a given channel as it is used by different speakers. These variations in average speech power stem in part from individual differences in the loudness with which people speak, and in part from differences in the telephone instruments and lines connected to the channel inputs. The average speech power in a telephone channel is usually measured and expressed in terms of a quantity called volume, defined as the ratio in db of S_{Si} to a fixed reference power level S_{ref}:

$$V_{Si} = 10 \log (S_{Si}/S_{ref}) \tag{216}$$

The second probability distribution (Fig. 16) describes how the single channel volumes V_{Si} differ from the volume of the average speaker, i.e., from the volume V_S corresponding to the mean speech power S_S of the distribution of S_{Si}. The distribution of the volume differences

$$V_{Si} - V_S = 10 \log (S_{Si}/S_S) \tag{217}$$

rather than the volumes V_{Si} is given because different investigators have used different values for S_{ref} in Eq. (216). Moreover, measurements of single channel volumes at the toll test boards of commercial telephone systems, at different times and in different countries,[28,29] have led to a number of different values of S_S, not only because speakers have tended to speak more softly as the noise performance of telephone equipment has been improved, but also because the instruments used for measuring volume are themselves not especially well suited to the unambiguous determination of average speech power.[30] The important fact about these measurements of single channel speech volume, however, is that they all lead to distributions which conform rather closely to the normal distribution shown in Fig. 16.

In view of the relation between volume and average speech power, it follows that the single channel speech powers S_{Si} obey a log-normal distribution. Note however that the mean $\overline{V}_{Si}$ of the volume distribution does not equal the volume V_S corresponding to the mean S_S of the distribution of average speech power. In general, the difference between these two volumes is given by

$$\overline{V}_{Si} - V_S = - 0.115 \sigma_V^2 \tag{218}$$

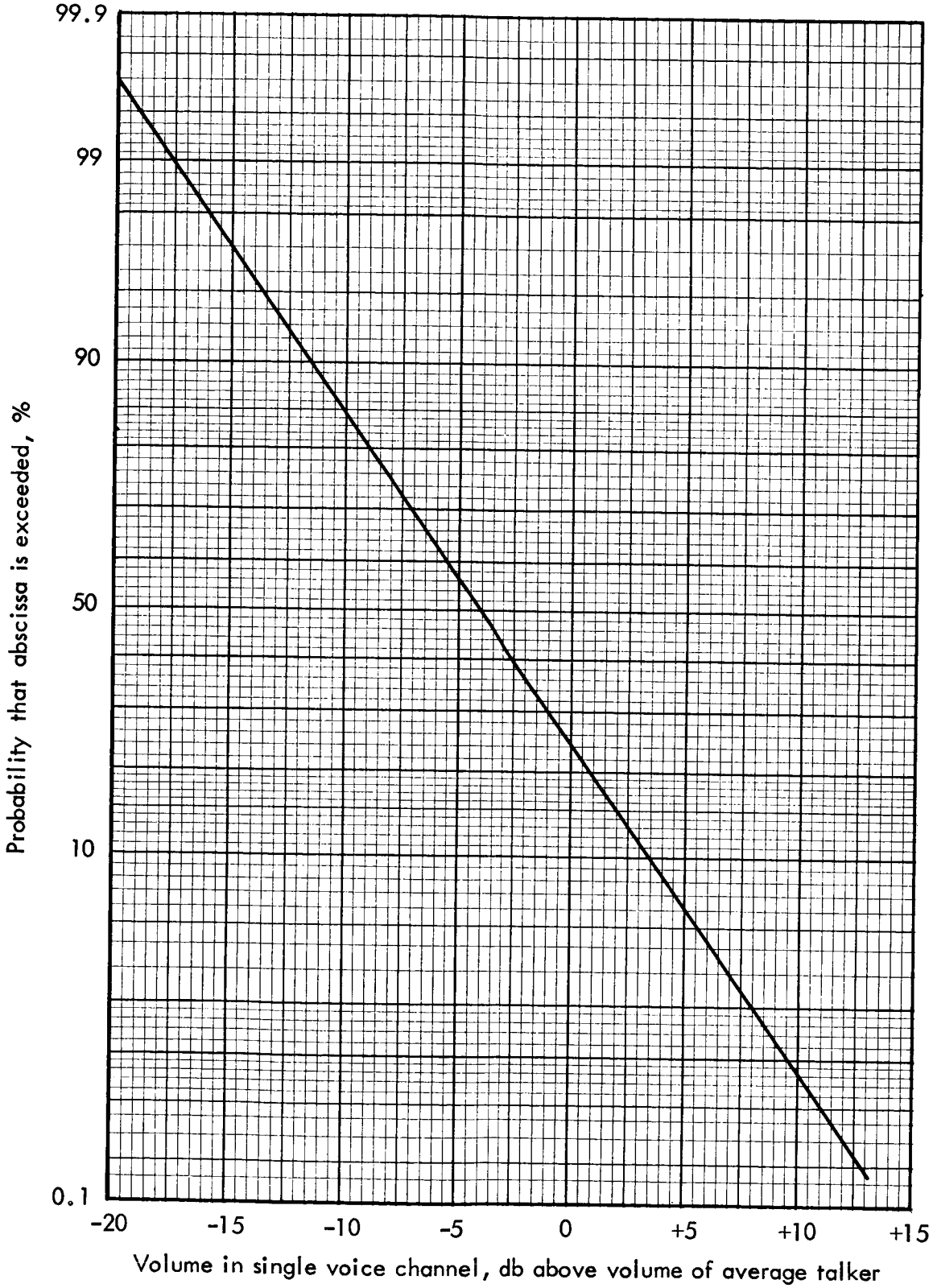


Fig. 16 — Distribution of volume for single voice channels

where σ_V is the standard derivation in db of the volume distribution. In the case of Fig. 16, σ_V = 5.8 db and $\bar{V}_{Si} - V_S$ = -3.8 db.

Definition of Peak Baseband Load*

The quantity of chief interest in connection with the multichannel speech baseband is the peak power or load it imposes on the system. Since the baseband signal, like the speech signals from which it is formed, can only be described statistically, the peak load is not the absolute maximum instantaneous power reached by the baseband. Rather, it is the power level which the baseband exceeds only for a specified small fraction of the time. This fraction will be called the overload probability** because it is the probability that the instantaneous baseband power overloads the system, i.e., exceeds the peak baseband load. The criterion for specifying the overload probability is that, with amplifiers just capable of handling the peak load without excessive distortion, the effects of the overloads that do occur should not be unacceptable to system users. Experiments with a number of baseband amplifiers in commercial telephone systems suggest that an overload probability of 0.1 percent is acceptable.

This would be a sufficient definition of peak baseband load if the speech volume in each of the individual channels were regulated to the same level and if the number of simultaneously active channels remained constant. In this case the distribution of the instantaneous baseband power would be stationary. It would be the same regardless

---

*Called "load capacity requirement" in Ref. 23.

**The term "overload expectation" sometimes used for this quantity is misleading since it is not an expectation in the statistical sense of the term.

of the time at which it was measured, and there would be no ambiguity about the level which the baseband exceeded 0.1 percent of the time. It would only be necessary to design the amplifier so that its overload threshold was equal to this level, and overloading would never be objectionable.

In practice, however, the number of simultaneously active channels changes continuously as speakers start and stop talking in the various channels of the system. At any instant, the number of active channels may range from 0 to Q where Q is the total number of channels in the circuit. In addition, the channel volumes are distributed over a wide range as indicated in Fig. 16. As a result, the actual baseband signal cannot be characterized by a stationary amplitude distribution.

In the face of these facts, a more comprehensive criterion of overload is required. Toward this end, note first that the changes in the number of active channels and in the volume in a given channel are much slower than the variations in the instantaneous baseband signal resulting from the rapid variations of the single channel speech waveforms. Thus a long sample of the baseband signal may be divided into a sequence of short intervals, defined by the condition that during each interval, both the number of active channels and the volumes in each of these channels are constant. Over each such interval, the distribution of instantaneous baseband power is well defined. By comparing the 0.1 percent level of this distribution with the amplifier overload threshold, it is possible to determine whether or not the amplifier was overloaded during that interval. In these terms, the generally accepted criterion for permissible

overloading on a commercial telephone system may be stated as follows: During the busiest traffic hour, the combined duration of all such intervals of overload shall not exceed 1 percent of the hour, or 36 sec.

In the case of an FDM speech baseband, then, the "peak baseband load," is defined as the power level which the instantaneous baseband signal exceeds more than 0.1 percent of the time, only during short intervals that total no more than 1 percent of the busiest hour. If an amplifier has the capacity to handle this power level without overloading, it should provide satisfactory performance during the busiest traffic hour.

Calculation of Peak Baseband Load

In order to compute the peak baseband load just defined, it is necessary to know, in addition to the single channel statistics of Figs. 15 and 16, the probability Pr(q) that during the busiest hour, exactly q out of a total of Q channels will simultaneously be active. Since the channels are independent, this probability is given by the binomial distribution

$$Pr(q) = \frac{Q!}{q!\ (Q-q)!}\ \tau^q(1-\tau)^{Q-q} \tag{219}$$

where τ is the probability that a given channel is active during the busiest hour. The parameter τ is called the "activity coefficient," and is generally assigned the value 0.25, thus suggesting that on the average a channel is active for a total of only 15 minutes even during the busiest hour of the day.

From its definition it is clear that the peak baseband load can be written as the product

$$\hat{S}_B(Q) = S_1(Q)\Lambda_B(Q) \tag{220}$$

where $S_1(Q)$ is the power level exceeded by the average baseband power during one percent of the intervals of constant channel activity into which the busiest hour can be divided, and $\Lambda_B(Q)$ is the 0.1 percent peak-to-average power ratio when the average baseband power has the value $S_1(Q)$.

To calculate $S_1(Q)$, consider the subset of constant-activity, constant-volume intervals of the busiest hour during which the number of active channels has a particular value q. The distribution of average baseband powers that occurs during such an interval is conveniently expressed in terms of a quality called "equivalent volume" which is defined by analogy to the volume of a single channel as

$$V_B(q) = 10 \log \left(\frac{\Sigma S_{Si}}{S_{ref}}\right)$$

where the summation includes only the q channels that are active. Assuming that the single channel volumes are normally distributed as shown in Fig. 16, Holbrook and Dixon[23] used numerical integration to compute the conditional probability $P_{V_B}(V|q)$ that, for the subset of intervals during which q channels are active, the baseband equivalent volume $V_B(q)$ would exceed V. Their results are shown in Fig. 17 for q = 1, 4, 16, and 64 where, as in Fig. 16, volume is expressed relative to the volume V_S of the average speaker. These curves

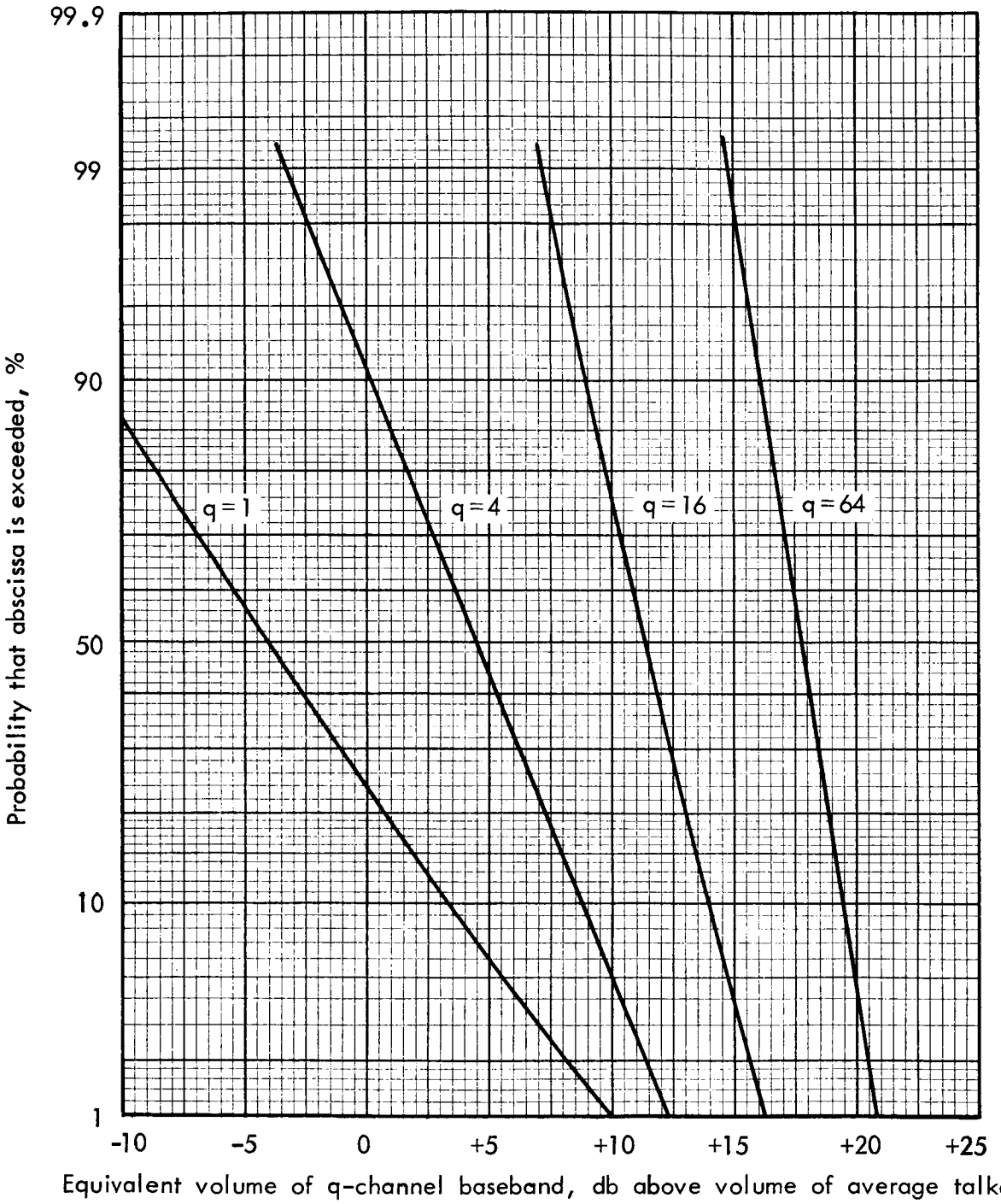


Fig.17-- Distribution of equivalent volume for basebands composed of q 100%-active voice channels with normally distributed volumes

illustrate quantitatively how, as the number of active channels increases, the fluctuations in equivalent volume due to differences in the volumes of the individual channels are reduced.

The distribution of baseband equivalent volume over all of the intervals that comprise the busiest hour--namely the total probability $P_{V_B}(V)$ that the Q-channel baseband equivalent volume exceeds V--may then be calculated from the summation:

$$P_{V_B}(V) = \sum_{q=1}^{Q} Pr(q)\ P_{V_B}(V|q) \tag{221}$$

where $P_{V_B}(V|q)$ is the probability given in Fig. 17 that, with q channels active, the equivalent volume exceeds V, and Pr(q) is the probability given in Eq. (219) that exactly q channels will be active. Assuming $\tau = 0.25$ in Eq. (219), the cumulative distributions for the equivalent volume of the Q-channel baseband are shown in Fig. 18 for Q = 3, 12, and 240.

If the equivalent volumes $V_1(Q) - V_S = 10 \log [S_1(Q)/S_S]$ exceeded 1 percent of the busiest hour are read from these curves and plotted as a function of Q, the curve shown in Fig. 19 is obtained. The desired values of the corresponding baseband power level $S_1(Q)$ may then be read directly from this graph once a value is assigned to S_S. Note however that the ratio $S_1(Q)/S_S$ affords a conservative design value for the baseband load factor Γ_i defined in Eq. (214) and appearing repeatedly in the SNR expressions of Table 7. It is the value of Γ_i that is exceeded only 1 percent of the busiest traffic hour.

To obtain a conservative approximation of the peak-to-average power ratio $\Lambda(Q)$ exceeded 0.1 percent of the time by the Q-channel

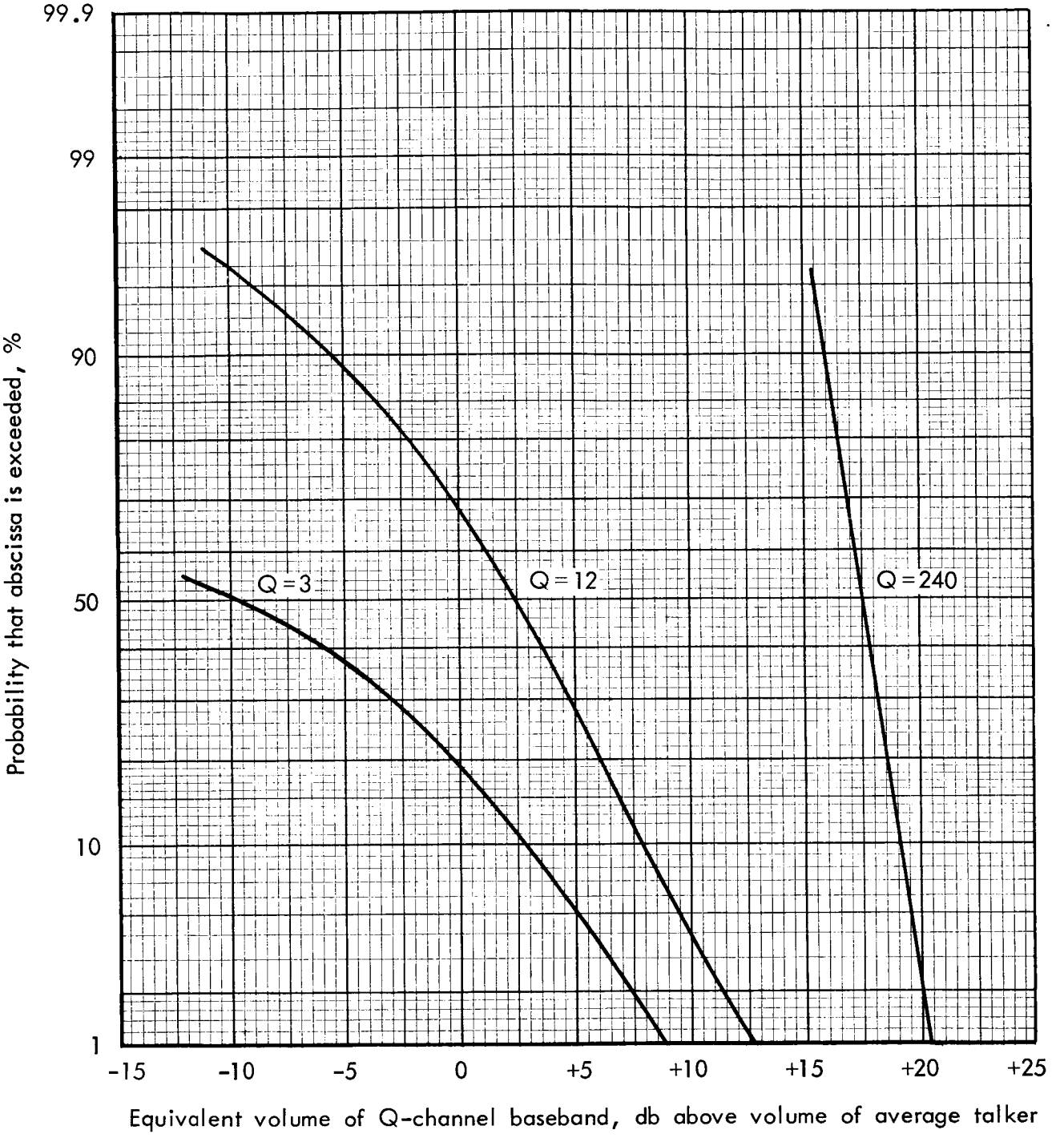


Fig. 18—Distribution of equivalent volume for basebands composed of Q 25%-active voice channels with normally distributed volumes

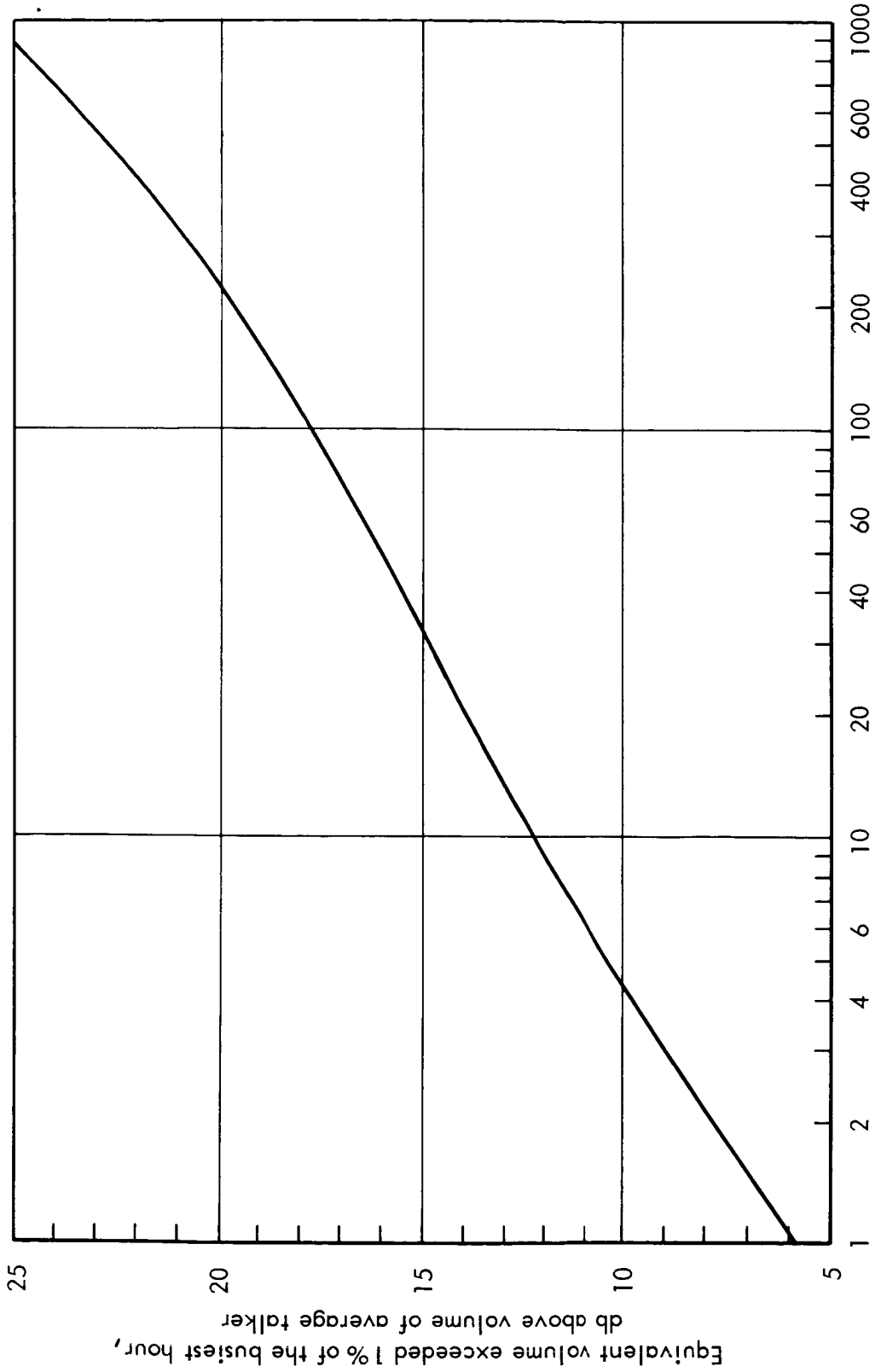


Fig. 19—Equivalent volume exceeded 1% of the busiest hour, for a baseband composed of Q 25%-active voice channels with normally distributed volumes

baseband when its average power is $S_1(Q)$, consider the simple but unrealistic case of a baseband formed from q 100 percent active channels where the volume in each channel is regulated to the same constant level. If this level is set equal to V_S, corresponding to the average talker, then the average power of the baseband is $\Sigma S_{Si} = q S_S$, and the equivalent volume is $V_B(q) = V_S + 10 \log q$. The distributions of the instantaneous baseband power about this average as determined experimentally by Holbrook and Dixon[23] for q = 4, 16, and 64, are given in Fig. 20.*

For comparison, Fig. 20 also includes the distribution for a single voice channel, for a gaussian-distributed signal, and for a pure sinusoid. It will be noted that, as the number of active channels increases, the baseband amplitude distribution rapidly approaches the gaussian. Indeed, when the number of channels is 64 or greater the amplitude distribution of an FDM speech baseband is similar to and quite well represented by band-limited white noise. This fact forms the basis for various practical baseband loading tests.[31,32,33]

From the curves of Fig. 20, the peak-to-average ratios may be determined as a function of q, not only for the ratio exceeded 0.1 percent of the time but for other overload probabilities as well. The result is plotted in Fig. 21, which includes, in addition to the peak-to-average power ratios exceeded 0.1, 1, and 10 percent of the time, a dashed curve representing what Holbrook and Dixon called the

*In these experiments, the individual channels were simply added together in the same audio frequency band rather than being frequency-division multiplexed. Holbrook and Dixon assert that any errors in the resultant amplitude distributions disappear as soon as the number of channels exceeds a few.

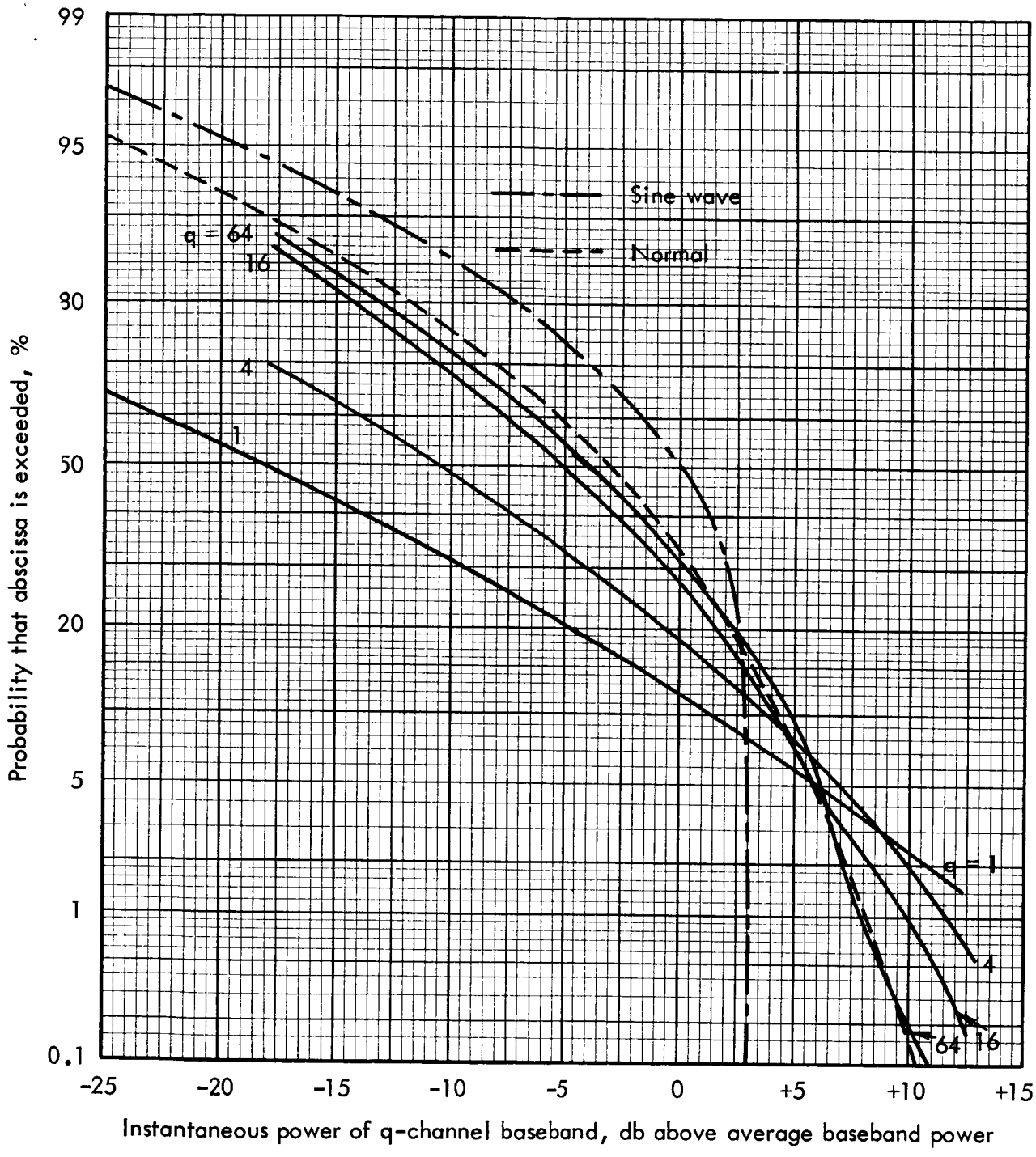


Fig. 20--Distribution of instantaneous power for basebands composed of q 100%-active voice channels with equal and constant volumes

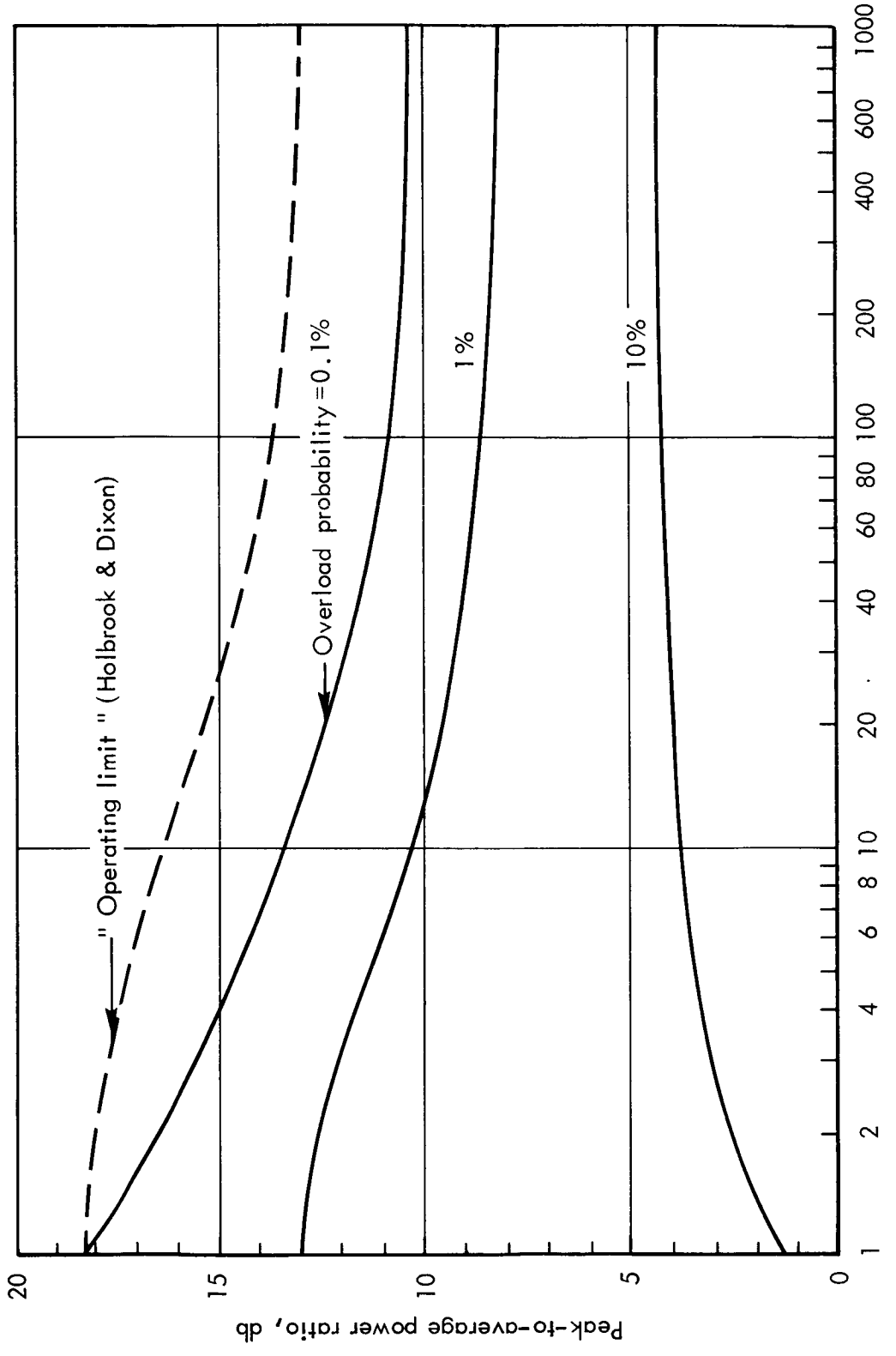


Fig. 21 — Peak-to-average power ratios exceeded with various probabilities for basebands composed of q 100%-active voice channels with equal and constant volumes

"multi-channel peak factor." This quantity is just the peak-to-average ratio for an unspecified overload probability of something less than 0.1 percent of the time and is therefore an even more stringent criterion of what constitutes an overload than has been adopted here. In any case it is obvious that the peak-to-average ratio decreases as the number of channels increases. It is also obvious that in the general case, where neither the number of active channels nor the channel volumes are controlled, the number of channels that are active when the baseband equivalent volume approaches its 1 percent level will nearly always be greater than the average number τQ of active channels. Therefore the 0.1 percent peak-to-average ratio corresponding to $q = \tau Q$ channels in Fig. 21 may be adopted as the desired approximation for $\Lambda_B(Q)$.

When values of $S_1(Q)/S_S$ obtained from Fig. 19 and estimates of $\Lambda_B(Q)$ based on Fig. 21 are substituted into Eq. (220), the values of peak baseband load plotted versus Q in Fig. 22 are obtained. Again the ordinate is actually the corresponding equivalent volume relative to the volume of a single speaker

$$V_B(Q) - V_S = 10 \log \left[\hat{S}_B(Q)/S_S\right]$$

which tells the number of decibels by which the peak baseband power exceeds the average speech power of the average talker.

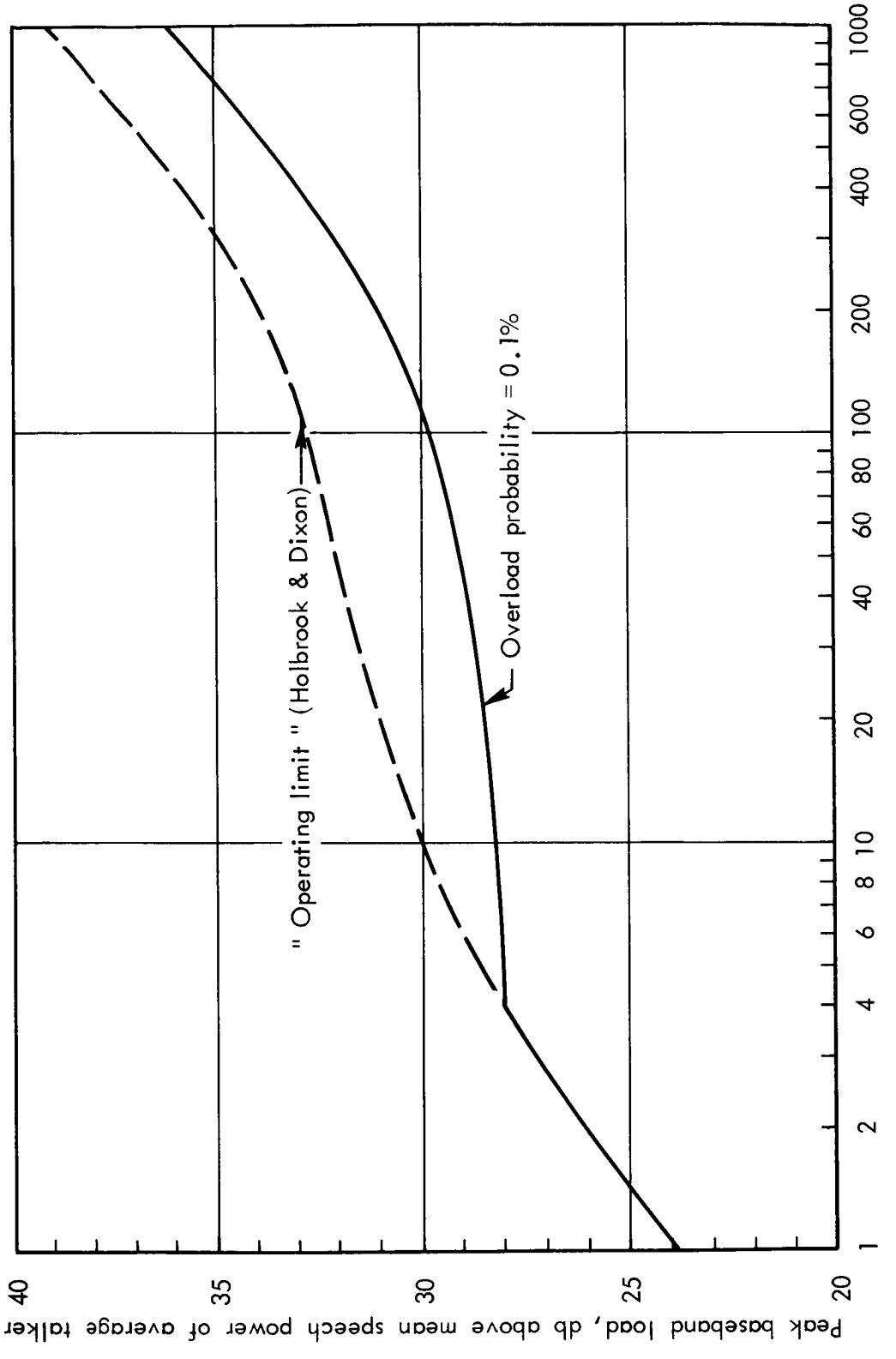


Fig. 22—Peak baseband load for FDM basebands composed of Q 25%-active voice channels with normally distributed volumes

IV. SIGNAL PROCESSING

INTRODUCTION

The general signal-to-noise expressions developed for several analog modulation methods in the preceding two sections can be used for comparing the "cost-effectiveness" of these methods in multichannel radio link applications. In this context, "effectiveness" is measured by the signal-to-noise ratios at the single channel outputs of the link and "cost" by the RF bandwidth and the peak or average power of the modulated carrier at the receiver input. For illustrative purposes, several types of signal inputs have been considered, including sinusoidal test tones, speech inputs, and random signals specified only by their amplitude distributions and power spectra. But in each case, no modification of signal characteristics was allowed except for those inherent to the modulation and multiplexing processes being compared. For example, the single channel inputs at the transmitting terminal were fed directly into the multiplexer, and its output, the baseband, was applied directly to the modulator. Similarly, at the receiver the demodulator output was fed directly to the demultiplexer whose signal outputs were just the desired replicas of the original input messages.

In the present section the constraint of no signal modification will be dropped, and new expressions will be developed to show how the cost effectiveness of various analog modulation methods can be improved by appropriate processing of the individual single channel inputs, the FDM baseband formed from them, or both.

As with modulation and multiplexing, the term "signal processing

technique" will normally denote a pair of complementary operations, in this case the input signal transformation applied prior to modulation at the transmitter and the inverse transformation following demodulation at the receiver. In normal practice, the latter operation undoes the effects on the signal of the former operation.

In most cases, the possibility of improving the cost-effectiveness of a modulation method by such techniques arises from the fact that the processing at the receiver is applied to both the signal and the noise, whereas the processing at the transmitter affects only the signal. Thus while the receiver processing merely restores the output signal to the form it would have had in an unprocessed system, it changes the properties of the output noise significantly. Indeed, if done properly, the processing at the receiver will discriminate against the noise, and for a given input signal power, will enhance the output signal-to-noise ratio.

Because the processing at the receiver is usually just the inverse of that at the transmitter, a signal processing technique can be identified by describing only the latter transformation; for this purpose it is sufficient to tell which signals are transformed and in what way the transformation changes them. In these terms, the techniques to be described in this section are:

Preemphasis. The signal is the baseband and the processing consists of modifying its spectrum in a specified way. The inverse transformation is called "deemphasis."

Compression. The signals are the single channel message inputs, and processing consists of reducing their dynamic range by

passing them through variable gain amplifiers.* The inverse transformation is called "expansion." Taken together, the pair of transformations is called "companding."

TASI (Time Assignment Speech Interpolation). The signals are those contained in a large number of telephone channels, and processing consists of selecting and multiplexing only the channels that are active. Since speech channels are active only a fraction of the time, the number of signals comprising the baseband is smaller than the number of channels from which they are selected; hence the baseband bandwidth is correspondingly reduced. It should be noted that TASI is not a signal processing technique in the same sense as preemphasis and companding since the signals involved are not intentionally modified. It might better be described as a signal switching technique for increasing the activity factor of the channels which comprise the baseband.

PREEMPHASIS

As just noted, the term preemphasis can be applied to any technique which modifies the power spectrum of the baseband signal. Normally this modification is produced by passing the baseband through a linear filter with an amplitude response that increases

*As explained later, the amplifiers in question are instantaneously linear with gains controlled by the input power averaged over a fraction of a syllable. Such "syllabic compression" is to be distinguished from the nonlinear "instantaneous compression" discussed in Ref. 2.

with increasing frequency. However, as will shortly be demonstrated, best results (i.e., greatest reduction in RF power or bandwidth) are obtained when the filter response is tailored in a fashion which depends on the shape of both the baseband power spectrum and the noise power spectrum at the demodulator output.

To determine the effect of arbitrary preemphasis, and, from it, the optimum filter response, recall from Table 5 that, for all modulation methods, S_R, the average power of the modulated carrier at the receiver input is related to K_L, the SNR at the demodulator output, by an expression of the form

$$S_R = CK_L \tag{224}$$

where the factor of proportionality C depends on the modulation/demodulation method and, for some methods, on the baseband peak-to-average power ratio as well. Specifically, C is given by:

	AM	DSB	SSB	PM	FM	SSBFM
$\frac{C}{N_o B_B} =$	$1 + \frac{\Lambda_B}{M^2}$	1	1	Λ_B/Φ^2	$\Lambda_B/3D^2$	$\sim 2\Lambda_B/9D^2$

(225)

where M, Φ, and D are respectively the modulation indices for AM, PM, and FM.

The same relation between the received signal power and the demodulator output SNR holds in a system where the baseband has been preemphasized. When applied to a preemphasized system, however, a superscript P is appended to each symbol to allow for the different value of K_L that will result from either a change in S_R or from a change in C

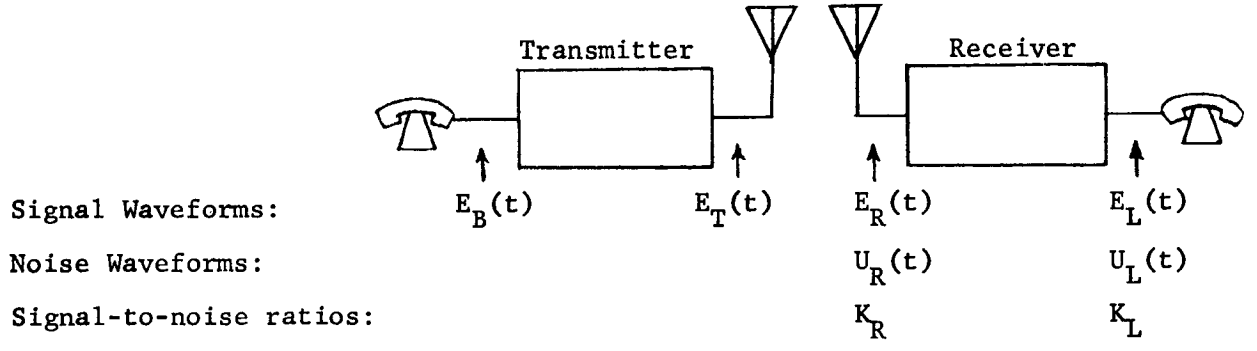
caused by the use of a different modulation index or by the effects of preemphasis on Λ_B. Thus,

$$S_R^P = C^P K_L^P \tag{226}$$

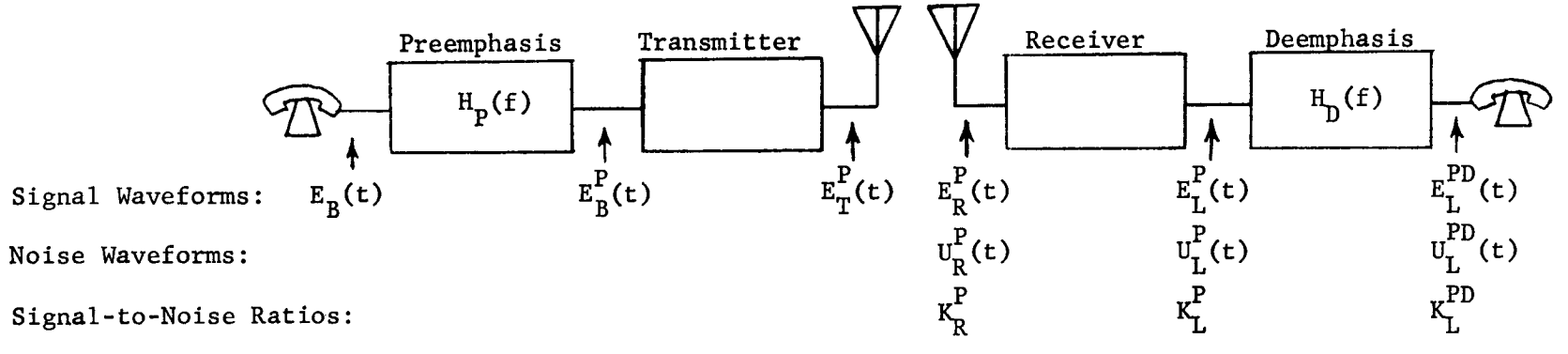
Single Channel Links

Equations (224) and (226) are perfectly general relations applicable equally to a single channel or to a multichannel baseband. The application to a single channel will be considered first--both because of its intrinsic practical importance (e.g., preemphasis is used routinely on commercial FM broadcasts) and its greater simplicity. Using Eq. (224) for a single channel link without preemphasis and Eq. (226) for the same link with preemphasis (same baseband, same receiver noise density, same propagation path, same modulation method, but possibly a different modulation index and transmitter power-antenna gain product), the problem is to determine, for the same output SNR, the ratio S_R^P/S_R of the received power required with preemphasis to that without. If this ratio is less than unity, then preemphasis has improved the cost-effectiveness of the link: it has reduced the power that must be received for a given link effectiveness. Figure 23 provides a functional block diagram of the two links being compared and illustrates the notation.

For the conventional link, the quality of the output signal is given directly by K_L in Eq. (224). For the preemphasized link on the other hand, output signal quality is given not by the demodulator output SNR, K_L^P in Eq. (226), but rather by K_L^{PD}, the SNR at the output of the deemphasis network that follows the demodulator. If the ratio of these two SNRs is ρ--i.e., if



(a) No Preemphasis



(b) With Preemphasis

Fig. 23—Block diagrams and notation for single channel link comparison

$$\rho \equiv K_L^P / K_L^{PD} \tag{227}$$

then the ratio of the received power required to produce the SNR K_L^{PD} with preemphasis to that which produces the SNR K_L without preemphasis is

$$\frac{S_R^P}{S_R} = \frac{C^P}{C} \rho \frac{K_L^{PD}}{K_L} \tag{228}$$

The desired figure of merit for preemphasis is then just the value I_P of this ratio that makes $K_L^{PD}/K_L = 1$

$$I_P \equiv \left(\frac{S_R^P}{S_R}\right)_{K_L^{PD} = K_L} = \rho \frac{C^P}{C} \tag{229}$$

Here, the factor ρ describes the net effect of the deemphasis network on the output SNR, and the ratio C^P/C, evaluated for the modulation method in question, takes into account the choice of modulation index and the effect of preemphasis on the baseband peak-to-average ratio.

Expressions for C^P and C may be read from Eq. (225); it remains to determine how ρ depends on the characteristics of the output signal and noise and on the choice of preemphasis and deemphasis networks. Suppose that on the conventional link, the power spectra of the signal and noise at the demodulator output are respectively $W_{E_L}(f) = C_1 W_{E_B}(f)$ and $W_{U_L}(f)$, where C_1 is a proportionality constant whose value depends on the modulation method and $W_{E_B}(f)$ is of course the baseband power spectrum. Then, in the preemphasized link, the corresponding spectra are

$$\begin{aligned} W_{E_L^P}(f) &= C_2 \left|H_P(f)\right|^2 W_{E_B}(f) \\ W_{U_L^P}(f) &= C_3 W_{U_L}(f) \end{aligned} \tag{230}$$

where C_2 and C_3 are additional constants of proportionality, and $H_p(f)$ is the transfer function of the preemphasis network. Hence the SNR at the demodulator output of the preemphasized link is

$$K_L^P \equiv \frac{S_L^P}{N_L^P} = \frac{\int_B W_{E_L}^P \, df}{\int_B W_{U_L}^P \, df} = \frac{C_2 \int_B |H_p|^2 W_{E_B} \, df}{C_3 \int_B W_{U_L} \, df} \tag{231}$$

where, for notational simplicity, the frequency dependence of the integrands is not explicitly indicated and the subscript on the integral sign means that the integration need cover only the baseband, $|f| \leq B_B$, since the integrands vanish outside this frequency range. In a similar way, if $H_D(f) = 1/H_p(f)$ is the transfer function of the deemphasis net, then the SNR at its output is

$$K_L^{PD} = \frac{\int_B |H_D|^2 W_{E_L}^P \, df}{\int_B |H_D|^2 W_{U_L}^P \, df} = \frac{C_2 \int_B W_{E_B} \, df}{C_3 \int_B |H_D|^2 W_{U_L} \, df} \tag{232}$$

where the second equality follows from Eq. (230) and the fact that $|H_D|^2 \cdot |H_p|^2 = 1$. The ratio of Eq. (231) to Eq. (232) then yields the desired general expression for ρ:

$$\rho \equiv \frac{K_L^P}{K_L^{PD}} = \frac{\int_B |H_p|^2 W_{E_B} \, df}{\int_B W_{E_B} \, df} \cdot \frac{\int_B |H_D|^2 W_{U_L} \, df}{\int_B W_{U_L} \, df} \tag{233}$$

The appearance of this expression can be simplified by changing the variable of integration from f to $\nu = f/B_B$, introducing the normalized power spectra

$$w_S = w_S(\nu) \equiv W_{E_B}(\nu B_B) \Big/ \int_{-1}^{1} W_{E_B}(\nu B_B)\, d\nu \tag{234a}$$

$$w_N = w_N(\nu) \equiv W_{U_L}(\nu B_B) \Big/ \int_{-1}^{1} W_{U_L}(\nu B_B)\, d\nu \tag{234b}$$

and writing for the "power spectral response" of the preemphasis network

$$g_P = g_P(\nu) \equiv |H_P(\nu B_B)|^2 = |H_D(\nu B_B)|^{-2} \tag{234c}$$

With these notational abbreviations, Eq. (233) becomes

$$\rho = \int_{-1}^{1} g_P\, w_S\, d\nu \cdot \int_{-1}^{1} g_P^{-1}\, w_N\, d\nu \tag{235}$$

This is the factor by which the deemphasis network changes the RF power required to produce a prescribed output SNR on a preemphasized single channel link. It is obvious that the value of ρ depends not only on the preemphasis characteristic specified by g_P, but also on the spectral characteristics of the baseband signal and the output noise given by w_S and w_N. For a particular w_S and w_N, the optimum preemphasis net is the one whose g_P makes ρ as small as possible.

As it happens, it is easy to express the minimum value of ρ in terms of $w_S(\nu)$ and $w_N(\nu)$ and to find the function $g_P(\nu)$ which leads to this minimum; that is, to find the optimum preemphasis characteristic for the given signal and noise spectra. It is only necessary to note that if $f(\nu)$ and $g(\nu)$ are arbitrary real functions, integrable-square over the range (a,b), then Schwartz' inequality shows that

$$\int_a^b f^2(v)\, dv \int_a^b g^2(v)\, dv \geq \left(\int_a^b f(v)\, g(v)\, dv \right)^2 \tag{236}$$

In particular, if a = -1, b = 1, and

$$f^2(v) = w_S(v)\, g_P(v)$$

$$g^2(v) = w_N(v)\, g_P^{-1}(v)$$

Schwartz' inequality becomes

$$\rho \geq \rho_\ell$$

where

$$\rho_\ell \equiv \left(\int_{-1}^{1} \sqrt{w_S(v)\, w_N(v)}\, dv \right)^2 \tag{237}$$

is the minimum value of ρ. By inspection of Eq. (235), the optimum preemphasis characteristic (the function $g_p(v)$ which makes $\rho = \rho_\ell$) is

$$g_P(v) = \sqrt{\frac{w_N(v)}{w_S(v)}} \qquad \begin{matrix}\text{Optimum} \\ \text{preemphasis} \\ \text{characteristic}\end{matrix} \tag{238a}$$

or, in terms of the preemphasis transfer function $H_P(f)$ and the power spectra themselves

$$|H_P(f)|^2 = \sqrt{\frac{W_{U_L}(f) \Big/ \int_B W_{U_L}(f)\, df}{W_{E_B}(f) \Big/ \int_B W_{E_B}(f)\, df}} \tag{238b}$$

With the aid of these expressions it is possible to draw some quite general conclusions about the efficacy of preemphasis as a signal processing technique under various circumstances. There are two cases of special interest.

Case 1: The output noise power spectrum is flat: $W_{U_L}(f) =$ constant, or $w_N(v) = \frac{1}{2}$. This is the situation for the linear modulation methods AM, DSB, and SSB, and also for PM. Hence, by Eq. (238b), the optimum preemphasis network for these methods has a transfer function whose magnitude is proportional to the inverse fourth root of the normalized baseband power spectrum

$$|H_P| = [2w_S(v)]^{-1/4} = \left[2B_B W_{E_B}(f) \Big/ \int_B W_{E_B}(f)\, df\right]^{-1/4} \qquad (239)$$

With such optimum preemphasis, ρ takes on the minimum value given by Eq. (237)

$$\rho_\ell = \frac{1}{2}\left[\int_{-1}^{1} \sqrt{w_S(v)}\, dv\right]^2 = \frac{\left[\int_B \sqrt{W_{E_B}(f)}\, df\right]^2}{2B_B \int_B W_{E_B}(f)\, df} \qquad (240)$$

This is the smallest value that ρ can have for the baseband described by $w_S(v)$. For some basebands, optimum preemphasis will yield a greater improvement (a lower value of ρ_ℓ) than for others. But, it is easy to show that, regardless of the nature of the baseband, ρ_ℓ will never be greater than unity; at worst, optimum preemphasis will leave the signal-to-noise ratio unchanged. Thus, applying Schwartz' inequality, Eq. (236), with $a = -1$, $b = 1$, $f(v) = \sqrt{w_S(v)/2}$, $g(v) = 1$, it follows that

$$\rho_\ell \le \int_{-1}^{1} w_S(v)\, dv \equiv 1$$

A quantitative feel for how ρ_ℓ depends on the shape of the baseband spectrum may be obtained from the examples shown in Table 8. It is seen that when the baseband spectrum matches the noise--i.e., when the baseband is flat--there is no improvement, and no reason to use preemphasis. However, the spectra of most single channel basebands are not flat but fall off at high frequencies. Thus optimum preemphasis usually does lead to a value of ρ less than rather than equal to unity.

The net preemphasis improvement, the fractional reduction in the required RF power, is then given by Eq. (229) with $\rho = \rho_\ell$

$$I_{P\ell} = \frac{C^P}{C}\,\rho_\ell$$

Here, C^P/C is given by Eq. (225), and ρ_ℓ by Eq. (240). For DSB and SSB, $C^P/C = 1$, and the improvement is equal to ρ_ℓ

$$I_{P\ell} = \rho_\ell \qquad \text{DSB, SSB}$$

For 100 percent modulated AM, $C^P/C = (1 + \Lambda_B^P)/(1 + \Lambda_B)$, and the net improvement factor is

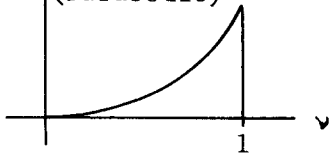
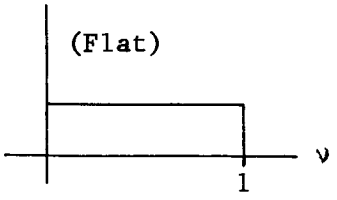
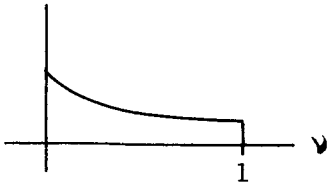
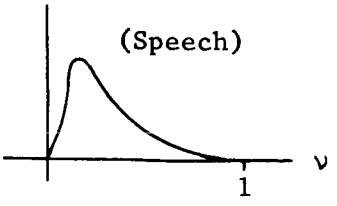
$$I_{P\ell} = \frac{1 + \Lambda_B^P}{1 + \Lambda_B}\,\rho_\ell \qquad \text{AM}$$

This will be smaller than ρ_ℓ if optimum preemphasis reduces the peak-to-average power ratio of the baseband. The same will be true for PM where

$$I_{P\ell} = \frac{\Lambda_B^P}{\Lambda_B}\left(\frac{\Phi}{\Phi^P}\right)^2 \rho_\ell \qquad \text{PM}$$

Table 8

IMPROVEMENT DUE TO OPTIMUM PREEMPHASIS ON SINGLE CHANNEL LINKS

Normalized baseband power spectrum $w_S(\nu)$	Reduction in required RF power due to optimum preemphasis of baseband $w_S(\nu)$	
	Case 1 Flat output noise power spectrum $w_N(\nu) = \frac{1}{2}$	Case 2 Parabolic output noise power spectrum $w_N(\nu) = \frac{3}{2}\nu^2$
$w_S = \frac{3}{2}\nu^2$ (Parabolic)  	$\rho_\ell = 0.75$ (-1.25 db)	$\rho_\ell = 1$ (0 db)
$w_S = \frac{1}{2}$ (Flat)  	$\rho_\ell = 1$ (0 db)	$\rho_\ell = 0.75$ (-1.25 db)
$w_S = \frac{2}{\pi}\frac{1}{1+\nu^2}$  	$\rho_\ell = .987$ (-0.1 db)	$\rho_\ell = 0.655$ (-1.85 db)
$w_S(\nu) = 1024\,\nu^2 e^{-16\nu}$ (Speech)  	$\rho_\ell = .497$ (-3 db)	$\rho_\ell = 0.091$ (-10.4 db)

If threshold considerations permit, it may be possible to reduce $I_{P\ell}$ still further by increasing the modulation index (the peak phase deviation Φ^P) of the preemphasized system.

Case 2. The output noise power spectrum is parabolic: $W_{U_L}(f) \propto f^2$, or $w_N(v) = \frac{3}{2} v^2$. This is the situation for the frequency modulation methods FM and SSBFM. Hence, by Eq. (238b), the optimum preemphasis for these methods has a transfer function

$$|H_P| = (\sqrt{3}\,|v|)^{1/2}\,[2w_S(v)]^{-1/4} = 1.316\sqrt{\frac{|f|}{B_B}}\left[\frac{2B_B W_{E_B}(f)}{\int_B W_{E_B}(f)\,df}\right]^{-1/4}$$

Comparing this with Eq. (239) it is seen that, for a given baseband, the magnitude of the transfer function for optimum preemphasis with FM differs from that for AM only by the factor $1.316\sqrt{|f|/B_B}$. With optimum preemphasis in this case, Eq. (237) gives

$$\rho_\ell = \frac{3}{2}\left[\int_{-1}^{1} |v|\sqrt{w_S(v)}\,dv\right]^2 = \frac{3\left[\int_B |f|\sqrt{W_{E_B}(f)}\,df\right]^2}{2B_B^2\int_B W_{E_B}(f)\,df} \qquad (241)$$

Again Schwartz' inequality can be applied to show that, with optimum preemphasis,

$$\rho_\ell \leq 1$$

As in Case 1, the magnitude of ρ_ℓ will be different for different baseband power spectra. Examples are given in Table 8, using the same normalized baseband spectra considered in connection with Case 1.

Here it should be noted that for a particular baseband spectrum, optimum preemphasis usually yields a greater improvement (lower value of ρ_ℓ) than in Case 1. Also note that, as in Case 1, the baseband for which there is no improvement ($\rho_\ell = 1$) is the one whose spectrum has the same shape as the noise.

This is no accident; it is another example of a general result that can be deduced directly from the definition of ρ_ℓ. Thus, whenever the baseband spectrum has the same shape as the output noise spectrum, when $w_S(\nu) = w_N(\nu)$, it follows from Eq. (237) that

$$\rho_\ell \equiv \left(\int_{-1}^{1} \sqrt{w_S(\nu)\ w_N(\nu)}\ d\nu \right)^2 = \left(\int_{-1}^{1} w_S(\nu) \right)^2 = 1$$

Equation (229) for the overall improvement due to optimum preemphasis also applies to Case 2, but now ρ_ℓ is given by Eq. (241), and the values of C^P/C appropriate to modulation methods that yield parabolic noise spectra must be used. In particular, for FM, Eq. (225) gives

$$I_{P\ell} = \frac{\Lambda_B^P}{\Lambda_B} \left(\frac{D}{D^P} \right)^2 \rho_\ell \qquad \text{FM} \qquad (242)$$

Thus the factor by which optimum preemphasis reduces the RF power required for a given output SNR depends not only on the baseband spectrum (which determines ρ_ℓ), but also on the effect of preemphasis on the baseband probability density function (which determines Λ_B^P/Λ_B), and finally on the modulation indices (D and D^P) used in the two cases.

The dependence of ρ_ℓ on the baseband spectrum has been discussed at some length, but not much in a general way can be said about the effect of preemphasis on the probability density function of an arbitrary baseband since the power spectrum and amplitude distribution are usually independent of one another. So far as the choice of the modulation indices D and D^P is concerned, there are two limitations to be considered. In each case the index must not exceed the smaller of two limiting values. One, D_a, is determined by the RF bandwidth B_{Ra} allocated to the link; the other, D_t, corresponds to threshold operation. For the link without preemphasis, these two values were given respectively by Eqs. (201d) and (201a):

$$D_a + 1 \cong B_{Ra}/2B_B \tag{243}$$

$$D_t^2\ (D_t + 1)^{3/2} = \Lambda_B\ K_L/(6\sqrt{10}) \tag{244}$$

For an optimally preemphasized link, the corresponding maximum indices are given by

$$D_a^P + 1 = B_{Ra}/2B_B \tag{245}$$

$$(D_t^P)^2\ (D_t^P + 1)^{3/2} = \Lambda_B^P\ K_L^P/(6\sqrt{10}) = \Lambda_B^P\ \rho_\ell\ K_L^{PD}/(6\sqrt{10}) \tag{246}$$

where, in the last equality, Eq. (227) was used to introduce K_L^{PD}, the output SNR after deemphasis. Note from Eqs. (243) and (245) that $D_a^P = D_a$.

Now, if RF bandwidth is the governing constraint both with and without preemphasis--if $D = D_a < D_t$ and $D^P = D_a^P < D_t^P$ --then $D = D^P$ and Eq. (242) reduces to

$$I_{P\ell} = \rho_\ell \frac{\Lambda_B^P}{\Lambda_B} \qquad \text{FM (Bandwidth-limited modulation indices)} \qquad (247)$$

On the other hand, if RF bandwidth is not limited by allocation in either case and if both links are operated at threshold ($D = D_t$, $D^P = D_t^P$), then D and D^P may be determined in terms of Λ_B, Λ_B^P, and ρ_ℓ from Eqs. (244) and (246) with K_L and K_L^{PD} set equal to the desired output SNR. If, in addition, D_t and D_t^P are considerably greater than unity, Eqs. (244) and (246) yield the approximation

$$\frac{D_t}{D_t^P} \cong \left(\frac{\Lambda_B}{\rho_\ell \Lambda_B^P} \right)^{2/7} \qquad D_t,\ D_t^P \gg 1$$

whence the improvement with optimum preemphasis is given by Eq. (242) as

$$I_{P\ell} = \left(\frac{\rho_\ell \Lambda_B^P}{\Lambda_B} \right)^{3/7} \qquad \text{FM (Threshold-limited modulation indices } D_t,\ D_t^P \gg 1) \qquad (248)$$

Since the bracketed quantity is normally less than unity, comparison of Eq. (247) with Eq. (248) shows that the preemphasis improvement will not be as large when modulation indices are threshold-limited as it is when they are bandwidth-limited. In this connection, however, it should be remembered that the RF power requirements for threshold operation are less to begin with than is the case for bandwidth-limited links.

Multichannel Links

In the case of a single channel, the baseband power spectrum and a desired output SNR were the given quantities. The question then was: What reshaping (preemphasis) of the baseband power spectrum will yield the greatest reduction in the RF power required to achieve the desired SNR? The answer was found to depend not only on the given baseband spectrum but also on the spectrum of the output noise, as determined by the modulation method used for transmission.

The corresponding problem for a multichannel circuit is the following: given the power spectra of a number of single channel message inputs combined by FDM into a baseband, and given the output SNRs desired for each, what modification of message spectra will produce the greatest reduction in the total RF power required to achieve the desired SNRs? It should be noted that neither the individual channel spectra, their bandwidths, nor the desired SNRs need have the same values for all channels, although in practice they usually do.

The first thing to be noted in approaching this problem is that, in contrast to the situation with a single channel baseband, it is not generally possible on an unpreemphasized link to exactly meet all of the SNR objectives for an arbitrary multichannel baseband. There is usually a "worst" channel, in the sense that if RF power is adjusted to make the output SNR in this channel just equal to the required value, the SNRs in the other channels will be at least equal to, and in many cases higher than, the specified values.

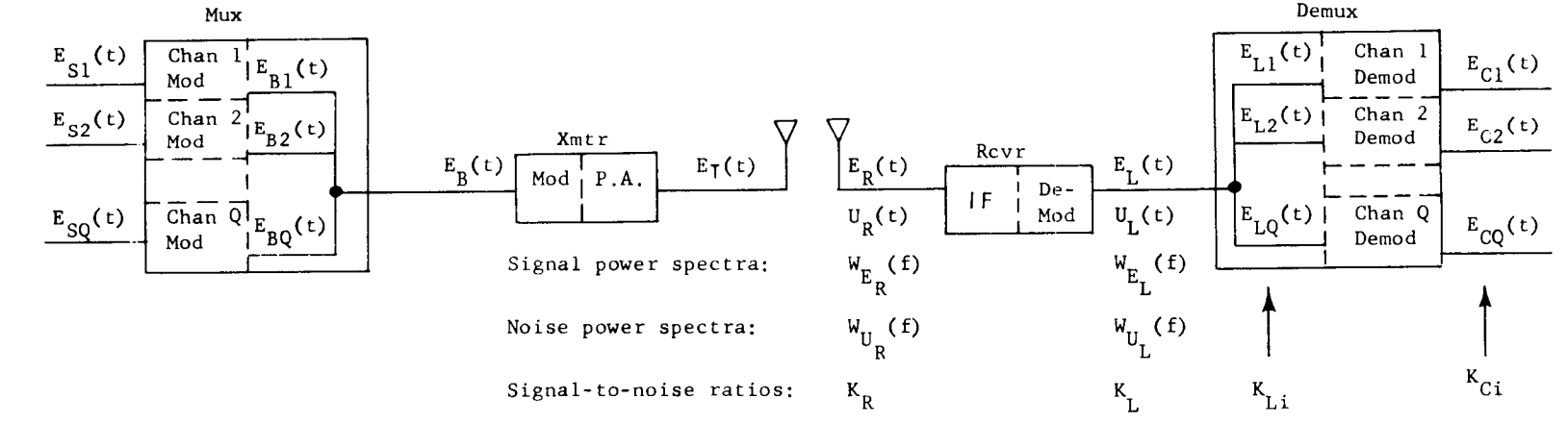
A simple and important practical example of this is a multichannel telephone link using FM. Here each of the single channel messages has

essentially the same bandwidth, spectrum, and SNR objective, but the parabolic character of the output noise power spectrum makes the single channel output SNRs grow progressively worse toward the high frequency end of the baseband (see Table 6).

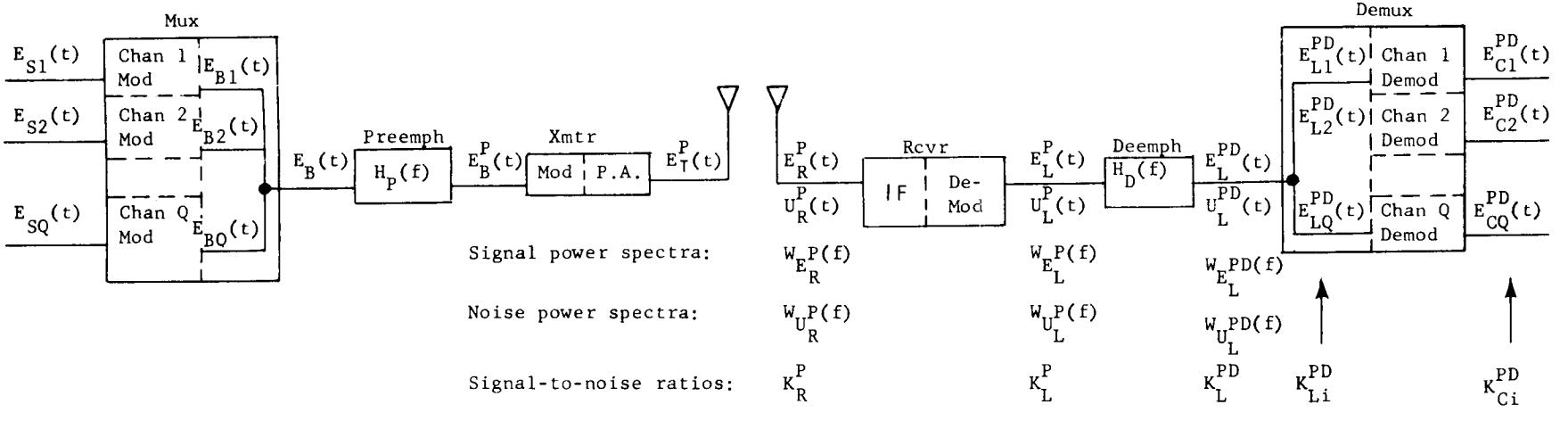
With preemphasis, on the other hand, each channel's share of the baseband power can be adjusted so that at the demultiplexer output the SNR in each channel will have the desired value. By eliminating the RF power wasted in producing unnecessarily high SNRs in the "better" channels, preemphasis reduces the RF power required to meet the specified SNR objectives.

Preemphasis in the multichannel case is commonly performed in the same way as with a single channel, that is, by passing the entire baseband through a preemphasis filter. However, it is also practical to preemphasize each individual channel prior to multiplexing. This can range from a simple level adjustment, in which the gains of the individual channel amplifiers in the multiplexer are adjusted to make the relative contribution of the different channels conform to a desired pattern, to a reshaping of each single channel spectrum by passing it through its own preemphasis filter. For purposes of analysis, however, the effects of both kinds of preemphasis will be represented by a single filter transfer function $H_p(f)$ operating only on the baseband waveform.

As in the single channel case, the magnitude of the preemphasis improvement will be measured by the ratio of the RF power with preemphasis to that without. A schematic block diagram of the two links to be compared is given in Fig. 24. For the link without preemphasis, the RF power that must be received in order to produce a SNR K_{Ci} at



(a) No Preemphasis



(b) With Preemphasis

Fig. 24—Block diagrams and notation for multichannel link comparison

the output of the i^{th} channel is

$$S_R = C\gamma_i K_{Ci} \tag{249}$$

where C is again given by Eq. (225), and from Table 7 the other factor of proportionality γ_i is given by

$$\frac{B_B}{B_{ci}}\gamma_i = \begin{cases} \Gamma_i & \text{AM, DSB, SSB, PM} \\ 3\left(\frac{f_i}{B_B}\right)^2 \Gamma_i & \text{FM, SSBFM} \quad B_{Ci} \ll f_i \end{cases} \tag{250}$$

Here, $\Gamma_i = S_B/S_{Bi}$ is the baseband load factor relative to the i^{th} channel, and f_i is the frequency of the i^{th} subcarrier.

If the specified output SNRs are K_i^o, i=1, 2, ···, Q, the power S_R required for the unpreemphasized Q-channel link is

$$S_R = \max\left\{C\gamma_1 K_1^o,\ C\gamma_2 K_2^o,\ \cdots,\ C\gamma_Q K_Q^o\right\} \tag{251}$$

i.e., S_R is the maximum of the set of powers calculated from Eq. (249) for i=1, ···, Q. If w (for "worst") is the number of the channel which requires this power to meet its SNR objective, then

$$S_R = C\gamma_w K_w^o \tag{252}$$

In the multichannel link with preemphasis, it is possible to choose the preemphasis transfer function $H_p(f)$ so that the channel

SNRs, K_{Ci}^{PD} (see Fig. 24b), exactly match the prescribed SNR objectives K_i^o. To see what condition this imposes on $H_P(f)$ requires an expression similar to Eq. (249) relating K_{Ci}^{PD} to the received signal power S_R. The needed relation can be obtained from Eq. (226) by expressing K_L^P, the SNR at the receiver output, in terms of the K_{Ci}^{PD}. For this purpose, recall that for each of the SSB channel demodulators in the demultiplexer, the output SNR is equal to the input SNR. Expressing the latter in terms of the signal and noise power spectra at the receiver output, the output SNR for the i^{th} channel is

$$K_{Ci}^{PD} = K_{Li}^{PD} = \int_i |H_D|^2 W_{E_L^P} \, df \Big/ \int_i |H_D|^2 W_{U_L^P} \, df$$

where the subscript i on the integral signs indicate that the integration covers only the i^{th} channel, and dependence of the integrands on f has been omitted for notational simplicity. The receiver output SNR can be expressed in a similar way

$$K_L^P = \int_B W_{E_L^P} \, df \Big/ \int_B W_{U_L^P} \, df$$

Combining these two equations and using Eq. (230) to express the power spectra in terms of the signal and noise spectra without preemphasis, the relation between K_L^P and K_{Ci}^{PD} is

$$K_L^P = \gamma_i^P \, K_{Ci}^{PD} \tag{253}$$

where

$$\gamma_i^P \equiv \frac{\int_i |H_D|^2 W_{U_L}\, df \Big/ \int_i W_{U_L}\, df}{\int_i W_{E_B}\, df \Big/ \int_B |H_P|^2 W_{E_B}\, df} \tag{254}$$

The preemphasis-dependent factor γ_i^P may be written more compactly in terms of the normalized signal and noise spectra introduced in Eqs. (234a)-(234c)

$$\gamma_i^P = \int_i g_P^{-1} w_N d\upsilon \cdot \int_{-1}^{1} g_P w_S d\upsilon \Big/ \int_i w_S d\upsilon \tag{255}$$

Finally, substituting Eq. (253) into Eq. (226), the desired relation between S_R and K_{Ci}^{PD} is obtained

$$S_R^P = C^P \gamma_i^P K_{Ci}^{PD} \tag{256}$$

This is a perfectly general relation in which the factor C^P depends on the modulation method, modulation index, and peak-to-average power ratio of the preemphasized baseband as shown in Eq. (225), and the factor γ_i^P depends on the power spectra of the signal and output noise, and on the preemphasis filter transfer function $H_P(f)$ as shown in Eq. (254). In order for a single value of RF power to yield the desired SNRs K_i^o for all channels, $H_P(f)$ must be chosen so that the γ_i^P satisfy the condition

$$\gamma_1^P K_1^o = \gamma_2^P K_2^o = \cdots\cdots = \gamma_Q^P K_Q^o = S_R^P / C^P \tag{257}$$

Using Eq. (255), and noting that

$$\int_i w_S d\nu = S_{Bi} / S_B = \Gamma_i^{-1} \tag{258}$$

this condition may be expressed explicitly in terms of $g_P = |H_P|^2$:

$$\Gamma_1 K_1^o \int_1 g_P^{-1} w_N d\nu = \Gamma_2 K_2^o \int_2 g_P^{-1} w_N d\nu = \ldots.$$

$$\ldots. = \Gamma_Q K_Q^o \int_Q g_P^{-1} w_N d\nu = \frac{S_R^P}{C^P \int_{-1}^{1} g_P w_S d\nu} \tag{259}$$

Given a preemphasis transfer function which meets this condition, S_R^P can be determined from any of the equations in Eq. (257) or (259). In particular for channel w

$$S_R^P = C^P \gamma_w^P K_w^o \tag{260}$$

Combining this with Eq. (252), the preemphasis improvement factor may be expressed in a form similar to Eq. (229) for single channel basebands

$$I_P = \frac{S_R^P}{S_R} = \rho_w \frac{C^P}{C} \qquad \text{Multichannel baseband} \tag{261}$$

where

$$\rho_w \equiv \gamma_w^P / \gamma_w \tag{262}$$

and γ_w and γ_w^P are given respectively by Eqs. (250) and (255).

Again there are two cases of practical importance distinguished by whether the output noise power spectrum is flat ($w_N = \frac{1}{2}$), or parabolic ($w_N = \frac{3}{2}\nu^2$). The expressions for ρ_w in these two cases are

$$\rho_w = \tfrac{1}{2} \frac{B_B}{B_{Cw}} \int_w g_P^{-1} d\nu \cdot \int_{-1}^{1} g_P w_S d\nu, \qquad \text{AM,DSB,SSB, PM} \tag{263}$$

$$\rho_w = \tfrac{1}{2} \frac{B_B}{B_{Cw}} \left(\frac{B_B}{f_w}\right)^2 \int_w \nu^2 g_P^{-1} d\nu \cdot \int_{-1}^{1} g_P w_S \, d\nu, \qquad \text{FM, SSBFM} \tag{264}$$

No systematic method has been developed for finding preemphasis transfer functions $H_p(f)$ which meet the general condition given by Eq. (259) in an optimum manner so that ρ_w will have its minimum value. However, in the commonly occurring special case of identical channels

$$\Gamma_i = \Gamma, \quad B_{Ci} = B_C = B_B / Q \tag{265}$$

and identical SNR objectives

$$K_i^o = K^o \tag{266}$$

it is easy to find a transfer function which satisfies condition (259) and can yield a significant reduction in ρ_w. Thus, substituting Eqs. (265) and (266) into Eq. (259) the condition on g_p is

$$\int_i g_p^{-1} w_N \, d\nu = \frac{S_R^P}{C^P K^o \Gamma \int_{-1}^{1} g_p w_S d\nu} \tag{267}$$

Since the right side of this equation is independent of i, g_p must be such that the integral on the left side has the same value for all channels. One obvious way of meeting this requirement is to choose

$$g_p(\nu) = w_N(\nu) = \begin{cases} \frac{1}{2} & \text{AM,DSB,SSB,PM} \\ \frac{3}{2}\nu^2 & \text{FM,SSBFM} \end{cases} \tag{268}$$

With this choice, Eqs. (263) and (264) yield

$$\rho_w = \begin{cases} 1, & \text{AM,DSB,SSB,PM} \\ \int_{-1}^{1} \nu^2 w_S(\nu) d\nu, & \text{FM, SSBFM} \end{cases} \tag{269}$$

where, in obtaining the second equality, it was noted that for FM, the worst channel is the one with the highest subcarrier frequency: $f_w = B_B$.

Thus with the simple preemphasis characteristic given by Eq. (268), there is no improvement in the case of the modulation methods which have a flat output noise power spectrum, whereas with the methods that have a parabolic output noise spectrum, there is an improvement whose magnitude depends on the baseband power spectrum $w_S(\nu)$. In particular for flat basebands ($w_S(\nu) = \frac{1}{2}$), Eq. (269) shows that $\rho_w = \frac{1}{3}$, corresponding to a 5 db reduction in S_R due to the effects of the deemphasis netalone. The net preemphasis improvement I_P is of course given by Eq. (261). Substituting the expressions for C^P and C from Eq. (225), the improvement with the preemphasis filter given by Eq. (268) is

$$I_P = \rho_w \frac{C^P}{C} = \begin{cases} 1 & \text{AM, DSB, SSB, PM} \\ \frac{1}{3} \frac{\Lambda_B^P}{\Lambda_B} \left(\frac{D}{D^P}\right)^2 & \text{FM, SSBFM} \end{cases} \tag{270}$$

No improvement was to be expected with the modulation methods having a flat output noise power spectrum, since there was in fact no preemphasis. With FM and SSBFM, however, it is seen that the net improvement depends in part on the effect of preemphasis on the baseband peak-to-average power ratio given by Λ_B^P. Since the preemphasis specified by Eq. (268) accentuates the channels near the high frequency end of the baseband, it can be expected that the baseband peak-to-average ratio for a given number of channels will in general have a value corresponding to some smaller number of channels. At the same time, Λ_B^P and Λ_B must have nearly the same value for a single channel baseband (Q = 1). In addition, they should approach the same asymptote for

large Q since the amplitude distribution of both the preemphasized and the unprocessed baseband must become more nearly gaussian as Q increases. Based on these arguments, the estimated dependence of Λ_B^P on Q is shown for speech basebands in Fig. 25.

The other factor to be considered when using Eq. (270) to compute the preemphasis improvement depends on the choice of modulation indices D and D^P for the two links being compared. As in the single channel case, these indices will be constrained by the threshold conditions given by Eqs. (244) and (246), and they may be restricted by bandwidth allocations as indicated in Eqs. (243) and (245).

The type of preemphasis described by Eq. (268) for multichannel FM links--namely

$$|H_p(f)| \propto f$$

is the only type normally employed on commercial telephone circuits. On microwave relay circuits, for example, the preemphasis characteristic is approximated by an RC filter whose amplitude response is nearly flat at the low frequency end of the baseband, and is so shaped that the bandwidth occupied by the modulated carrier is the same with preemphasis as without. It should be pointed out that this type of preemphasis amounts to little more than an adjustment of the relative levels of the individual channels that comprise the baseband. Moreover, the actual link improvement--when measured in terms of reduced RF power for a given minimum output SNR--amounts to only about 2 db.

It would appear that an additional improvement of 3 db or more could be obtained with all types of modulation by superimposing on the simple preemphasis characteristic of Eq. (268) a fine structure

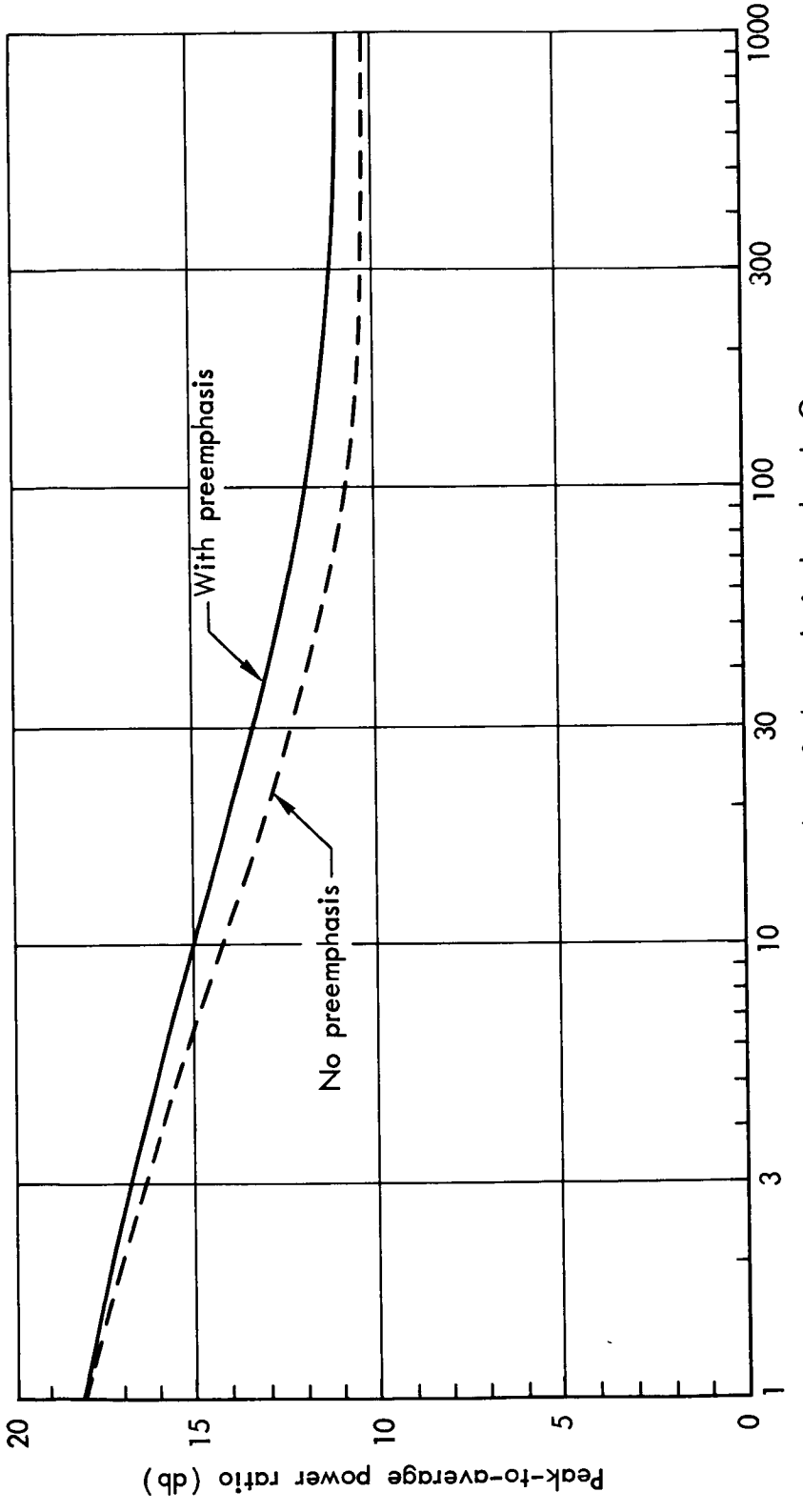


Fig. 25— Peak-to-average power ratio exceeded with 0.1% probability

that would correspond to preemphasizing each channel individually with the optimum preemphasis for SSB modulated single voice channels. The transfer functions of the required single channel filters would then depend on the single channel speech power spectrum as shown in Eq. (239). The estimate of a 3 db saving is based on the last entry of Table 8, which employs a simple analytical approximation to the normalized speech power spectrum. That this kind of preemphasis does not violate the general condition of Eq. (267) follows from the fact that it reduces each of the single channel integrals on the left side of this equation by the same factor.

COMPANDING

As previously noted, "companding" is a contraction of "compressing" and "expanding," a pair of complementary signal processing operations. Application of the former to the single channel input signals to the multiplexer and of the latter to the single channel outputs from the demultiplexer (see Fig. 1) can lead to significant improvements in the cost effectiveness of a multichannel link.[34,35]

To carry out these operations, each channel is equipped with a "compandor," consisting of a "compressor" at the channel input, and an "expandor" at the channel output. The compressor may be viewed as a variable-gain, instantaneously linear amplifier whose gain is controlled by the short-term average power at its input in such a way that strong signals are amplified less than weak ones, and very strong signals are attenuated. The expandor applies the inverse operation to the channel output signal. It may be viewed as a variable-loss attenuator whose loss is controlled by the short-term average power

at its input in such a way that strong signals are attenuated less than weak ones, and very strong signals are amplified. Extremely weak signals, such as noise in the absence of signal, are attenuated most of all.

Thus, the compressors reduce the dynamic range of the single channel inputs which form the baseband and the expandors restore their dynamic range after demultiplexing. The principal effects are first, to reduce the dynamic range of the baseband to permit more efficient transmission, and second, to modify the channel output signal and noise levels to yield higher output signal-to-noise ratios. To determine the magnitude of these effects, it is necessary to examine the operation of the compressor and expandor in greater detail. This will be done for each device in turn.

Compressor

The variable-gain amplifiers used to compress individual speech inputs that are to be frequency-division multiplexed are called syllabic compressors. This reflects the fact that the averaging time constant of their gain-controlling circuits is short (on the order of 3 msec) compared with the average length (250 msec) and build-up time (50 msec) of a typical syllable.* Thus the compressor gain variations will follow the envelopes of individual syllables. For example, accented syllables will be compressed more (i.e., the compressor gain will be lower) than unaccented or soft-spoken syllables.

---

*Actually, two time constants are involved.[34,36] The "attack time" is the time required for the compressor output signal envelope to reach a value 1.5 times its steady-state value following a sudden 12 db rise in input level. "Recovery time" is the time required for the output to reach .75 times its steady-state value following a sudden 12 db drop in input level. Nominal values are 3 msec for attack and 13.5 msec for recovery.

The manner in which compressor gain depends on input signal power is normally described in terms of the compressor's response to steady sinusoidal test tones. This response may be presented graphically in several equivalent ways, two of which are shown in Fig. 26. In both cases, the solid lines show how the long-term average output power depends on the long-term average input power for a steady 1000 cps tone. This dependence is characterized by three parameters:[36]

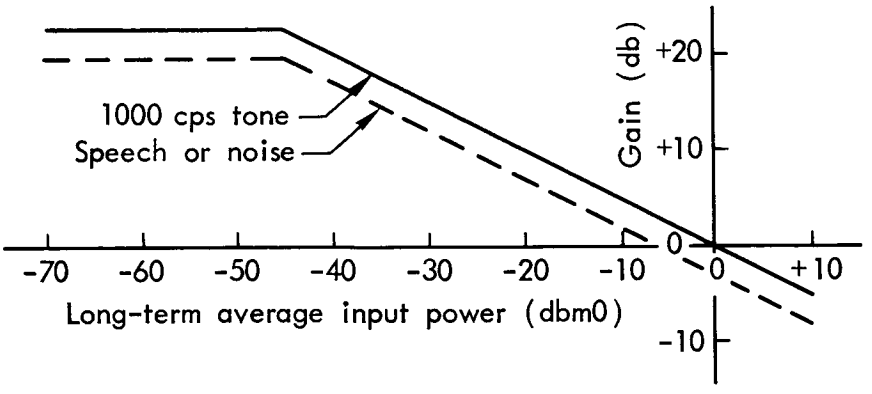
Crossover level or unaffected level. The average test tone input power for which the compressor gain is zero. Typical values in current use are 0 dbm0 (as shown in Fig. 26) and 5 dbm0.*

Compression ratio. The ratio of an input level in dbm0 to the difference between the corresponding output level and that corresponding to an input of 0 dbm 0. It is also the reciprocal of the slope of the curve shown in Fig. 26b. A compression ratio of 2 is used almost universally.

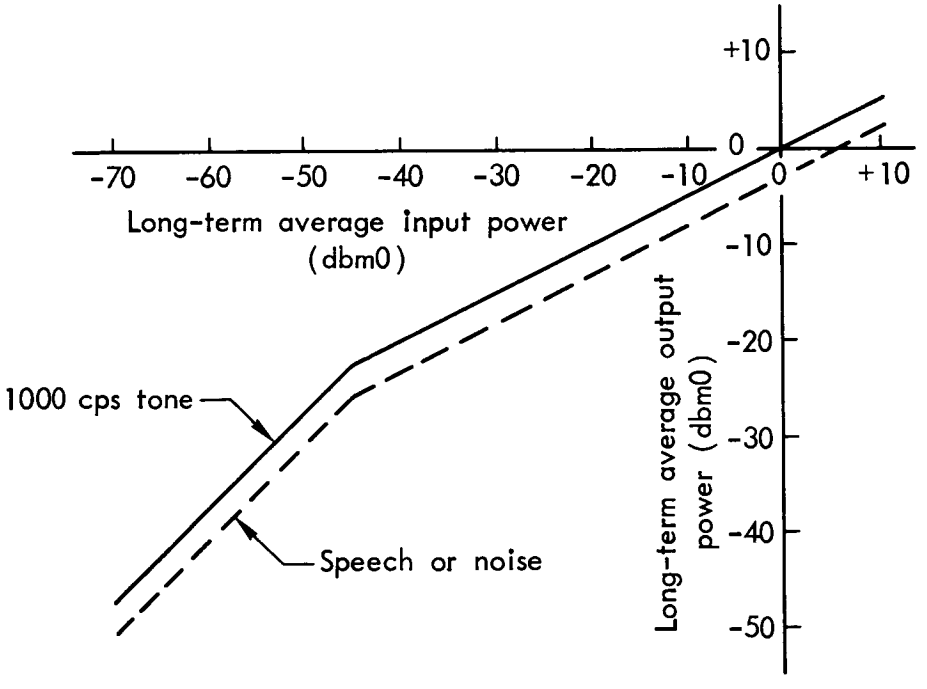
Range of level. The range of input powers in dbm0 over which the compression ratio has its prescribed value. A typical range of levels is from -45 dbm0 to +5 dbm0.

Unfortunately, this description of a compressor in terms of its steady-state response to sine waves is not sufficient to determine the effect of the compressor on a speech waveform. For this purpose, and also for describing the characteristics of an FDM baseband formed from compressed speech channels, it is convenient to separate the effect

*The unit "dbm0" implies that power is expressed in db relative to 1 milliwatt at a so-called point of "zero relative level"--i.e., the single channel input terminals or any other point at which the same signal levels are developed.



a. Gain versus input power



b. Output power versus input power

Fig. 26—Characteristics of a typical compressor

of a compressor on average speech power from its effect on instantaneous peaks in the speech waveform. As in Section III, the goal will be to describe the peak baseband load, which was there defined as the product of the average baseband power that is exceeded only one percent of the busiest hour and the peak-to-average power ratio when the average baseband power is at this level.

Consider average power first, and note that the average power at the compressor output depends not only on the average input power, but also on the nature of the input waveform. With a steady 1000 cps sine-wave input, the 3 millisecond average input power, and hence the compressor gain, is constant. With a speech waveform, however, the compressor gain changes continually in the fashion described earlier. In particular, experiments show that for the same long-term average input power, the average output power for both speech and noise-like waveforms is about 3 db lower than for a 1000 cps test tone. This result is plotted in Figs. 26a and 26b using dashed lines.

These curves show the effect of a compressor with a zero dbm0 crossover level on the average speech power in a single channel. In order to assess the effect of compression on the average power of a Q-channel FDM baseband, it is first necessary to determine its effect on the distribution of single channel speech powers at the multiplexer input. For this purpose, recall from Section III that the average speech power S_{Si} in an active channel may be expressed in terms of the volume difference

$$V_{Si} - V_S = 10 \log\left(S_{Si}/S_S\right)$$

defined in Eq. (217). Also recall that the volumes in the active channels follow the gaussian distribution shown in Fig. 16. With this volume distribution at the compressor inputs of a companded multichannel link, the volume distribution at the compressor outputs is easily found by increasing each volume in Fig. 16 by the gain indicated by the dashed line in Fig. 26a. With S_S = -9.9 dbm0, the result of doing this for compressors with crossover ratios of 0 dbm0 and 5 dbm0 is shown by the solid curves in Fig. 27. Observe that in both cases the volume distribution is still gaussian, but compared with the uncompressed volume distribution shown in Fig. 16, the standard deviation is cut in half.

Given these distributions of the volume of the compressed single channel speech inputs to the multiplexer, the level $S_1^C(Q)$ exceeded during 1 percent of the busiest hour by the average power of the compressed Q-channel baseband may be calculated by exactly the same procedure that was outlined in Section III for unprocessed speech channels. Rizzoni[35] has made these calculations for compressors with 0 dbm0 crossover levels. The equivalent volume distributions shown in Figs. 7 and 8 of his paper correspond respectively to the distributions of noncompressed baseband volume shown in Figs. 17 and 18 of this Memorandum. The resultant values of $S_1^C(Q)$ and the corresponding equivalent volume $V_1^C(Q)$ are plotted versus Q in Fig. 28 which is based on Rizzoni's Fig. 9. Note that compressor crossover levels of both 0 dbm0 and 5 dbm0 are shown. In using these curves to compare baseband load factors with those shown in Fig. 19 for unprocessed basebands, it is important to use the special scales in Fig. 28; these scales express the baseband equivalent volume relative to the volume of the average talker after compression by

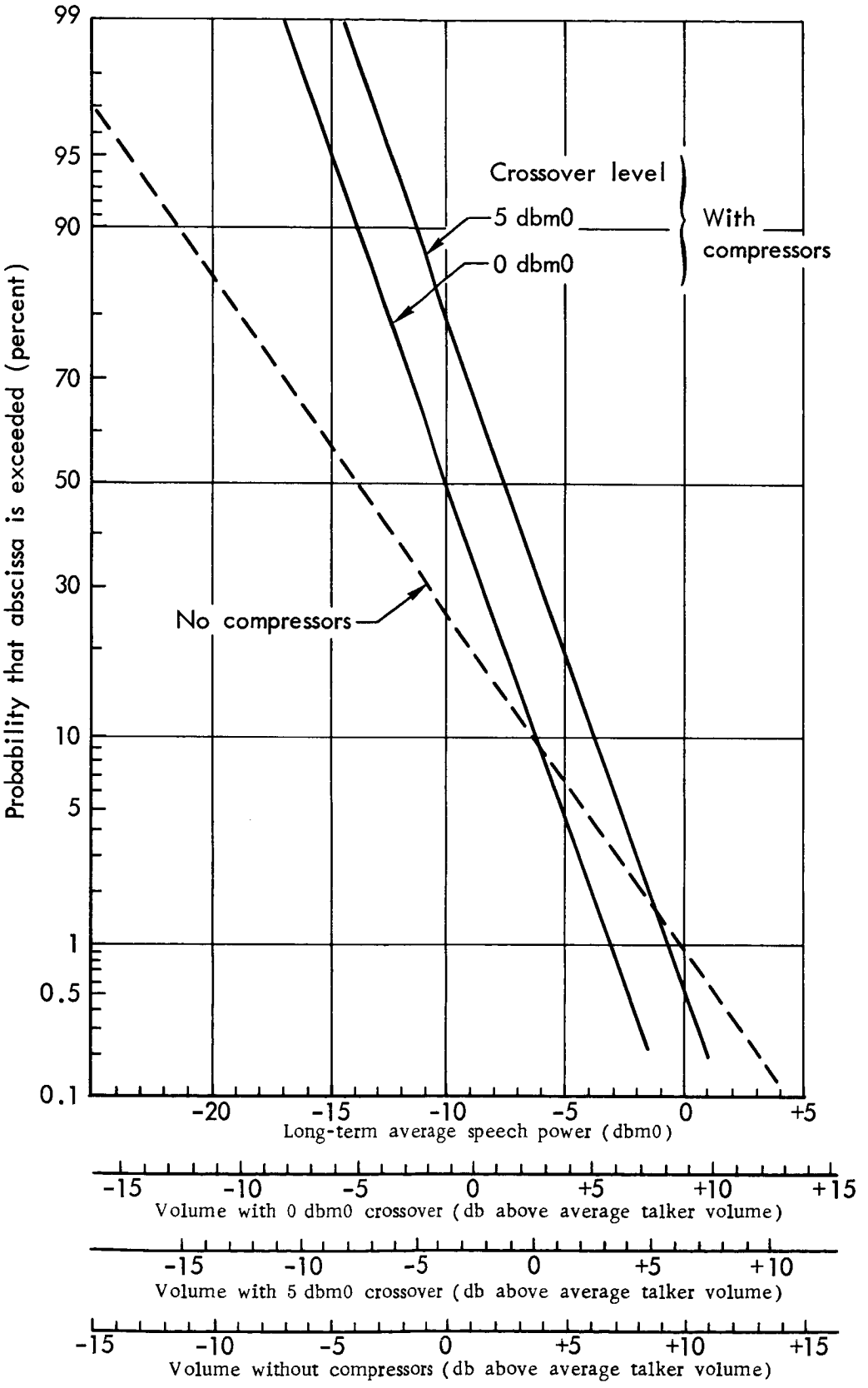


Fig. 27—Distribution of average single channel speech power and volume at multiplexer input

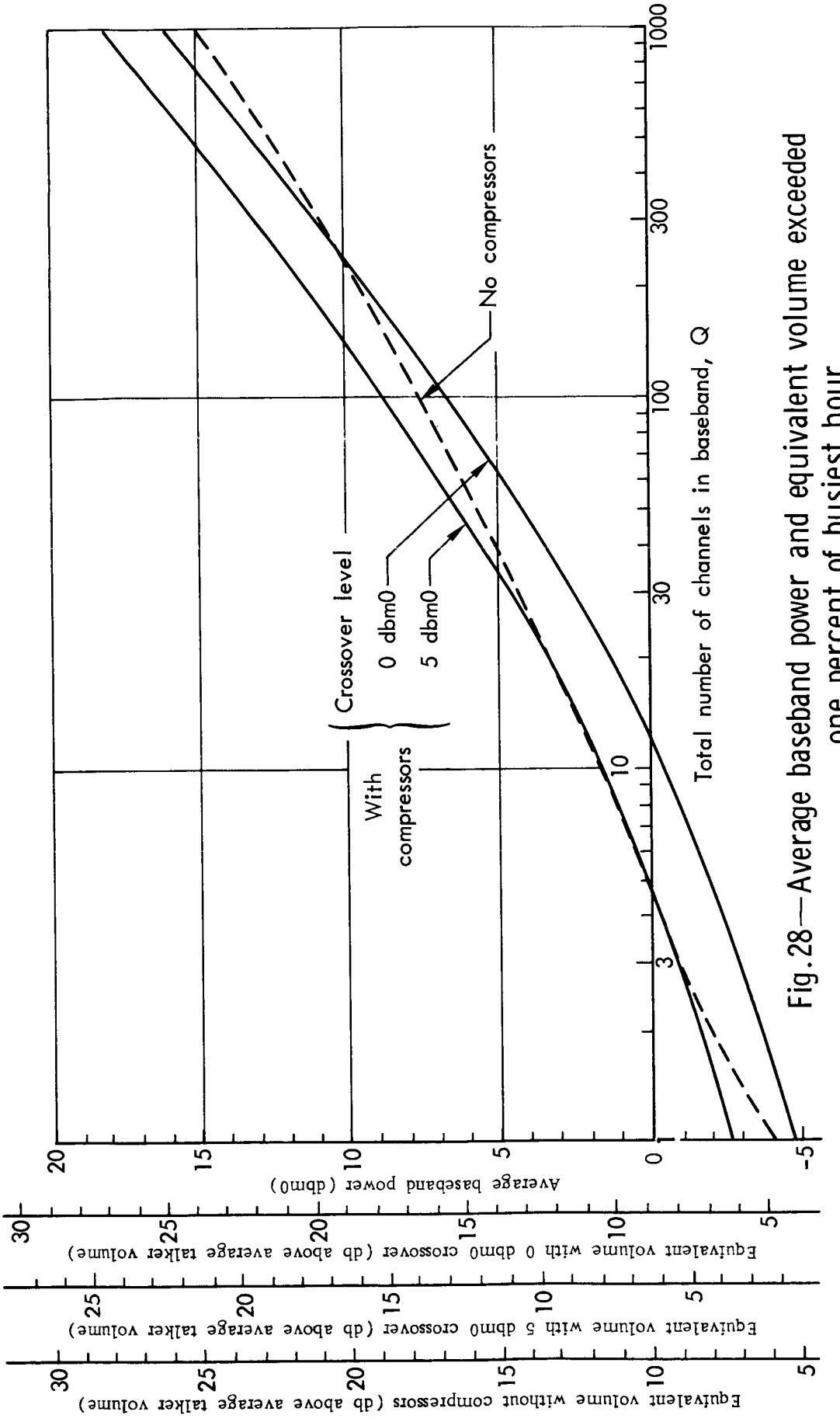


Fig. 28—Average baseband power and equivalent volume exceeded one percent of busiest hour

a compressor with the same crossover level as those used on the inputs to the baseband.

To find the peak baseband load, it remains to determine the effect of compression on the individual speech inputs and hence on the peak-to-average power ratio $\Lambda_B^C(Q)$ corresponding to the average baseband power $S_1^C(Q)$. While it can be argued on a qualitative basis that the majority of the instantaneous variations in the single channel inputs to the multiplexer will be compressed, and hence the baseband peaks reduced, this effect of compression is best determined experimentally. The results of such measurements[35] are shown in Fig. 29, where $\Lambda_B^C(Q)$ is plotted versus Q. Although the data were obtained for compressors with 0 dbm0 crossover level, they apply equally well to 5 dbm0 crossover level compressors. Just as with uncompanded channels, the distribution of instantaneous to average signal power is independent of the value of mean power or volume, which is the only baseband quantity that depends on the crossover level.

Finally, the values of $S_1^C(Q)$ from Fig. 28 are combined with the values of $\Lambda_B^C(Q)$ from Fig. 29 to obtain the peak power for the compressed baseband

$$\hat{S}_B^C(Q) = S_1^C(Q)\ \Lambda_B^C(Q)$$

A plot of $\hat{S}_B^C$ versus Q is given in Fig. 30 for both 0 and 5 dbm0 crossover level compressors.

Expandor

An expandor has already been described as a variable-loss attenuator whose loss is controlled by the short-term average power of the waveform

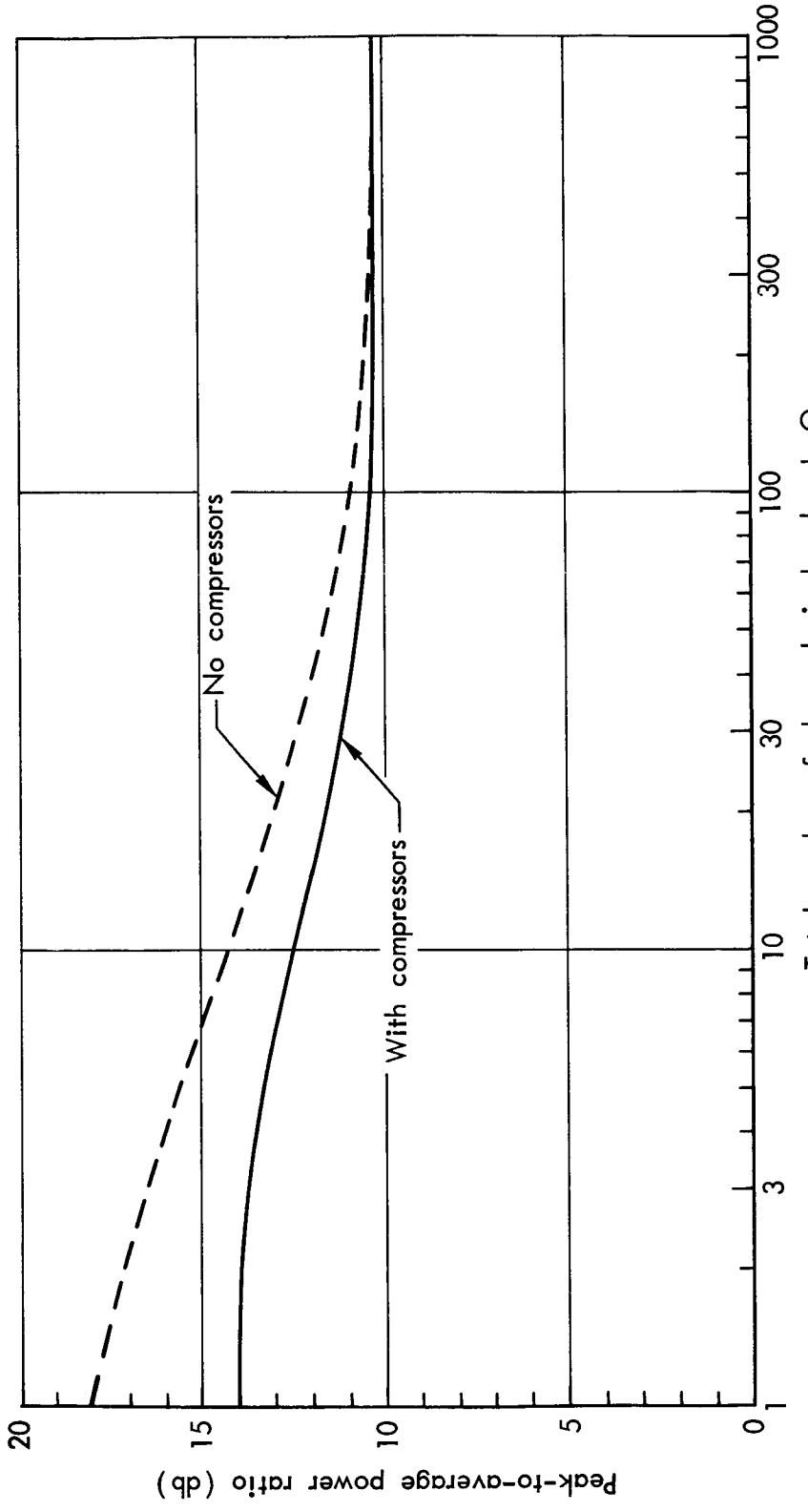


Fig. 29—Peak-to-average power ratio exceeded with 0.1% probability when baseband has average power shown in Fig. 28

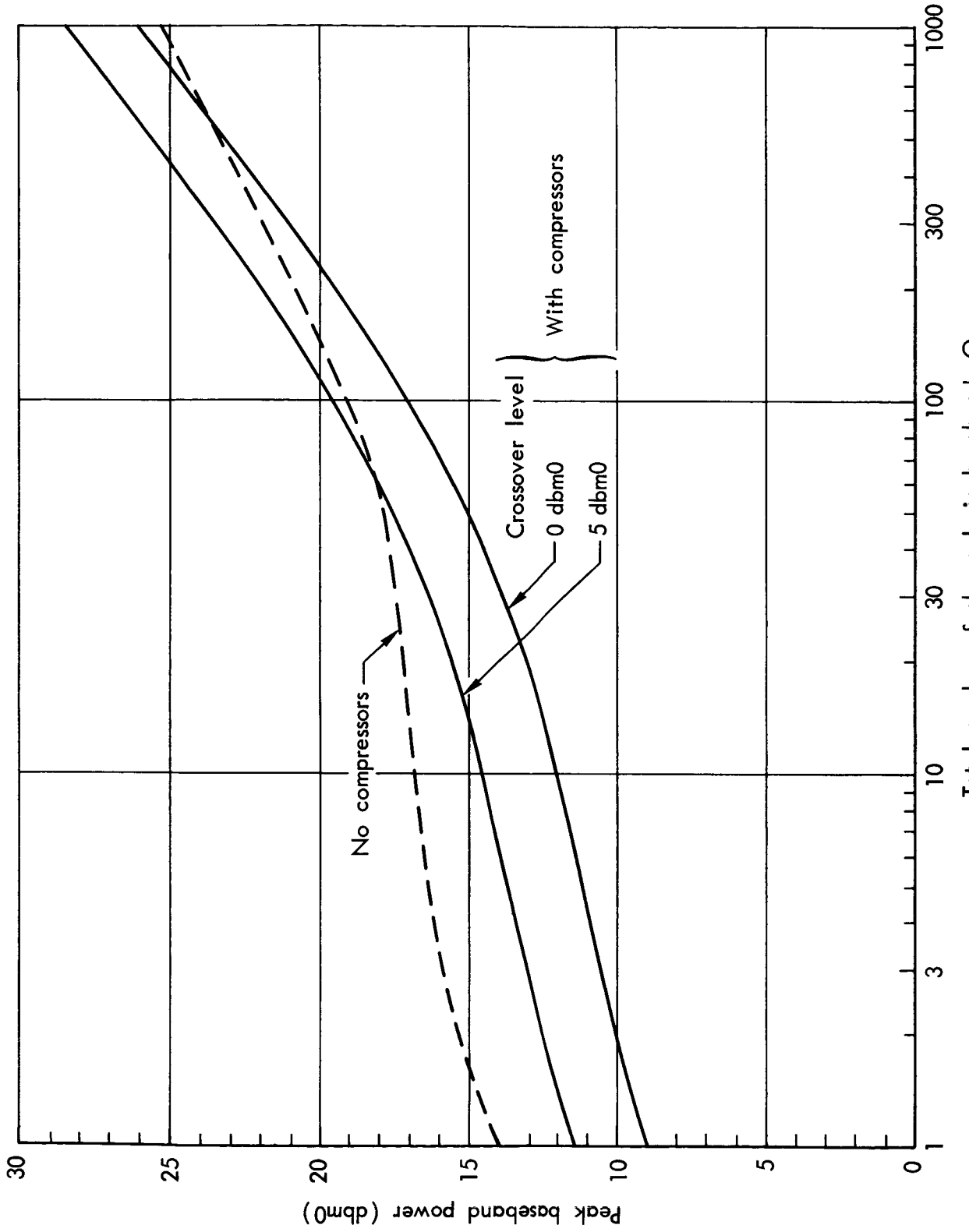
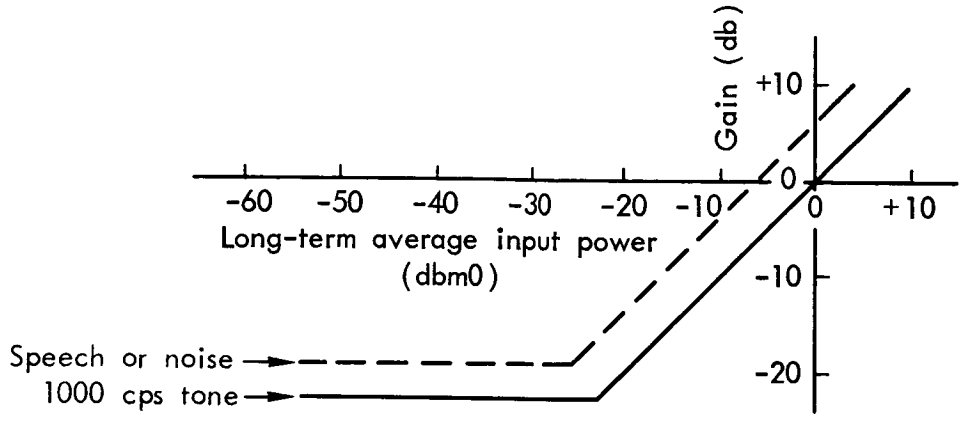


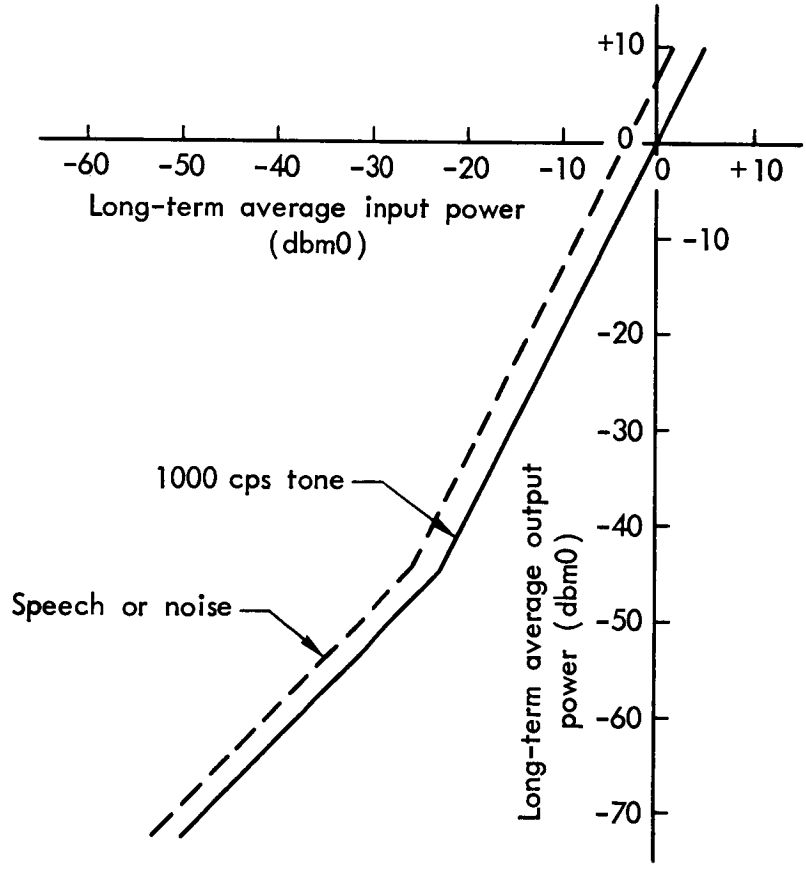
Fig. 30— Peak baseband load

at its input. Just as with its inverse, the compressor, the input-output characteristics of an expandor may be described in terms of its response to either steady test tones, or to speech and noise waveforms, as shown for example in Fig. 31. Also, by analogy with a compressor, the principal parameters of an expandor (in addition to its attack and recovery time constants) are its crossover level (the sine-wave input power at which the expandor loss is zero), its expansion ratio (the slope of the output versus input curve of Fig. 31b) and the range of sine-wave input levels over which the expansion ratio is constant. Thus, the expandor described in Fig. 31 has a 0 dbm0 crossover level, an expansion ratio of 2, and a range of input levels extending from -23 dbm0 to +5 dbm0.

In contrast to the situation with a compressor, the gain-controlling waveform at the input to an expandor is a mixture of the compressed single channel signal and the receiver output noise contained in the channel. At times when the channel is not active, i.e., not carrying speech power--the expandor loss is of course controlled by the noise level alone. On the other hand, when the channel is active, the short-term average speech power is normally much greater than the noise, and so controls the expandor loss. Thus the attenuation of the expandor is much higher in the absence of speech, and so the magnitude of the noise component at the expandor output is considerably smaller without speech than when speech is present. Extensive listening tests have shown that such a quieting effect during pauses and between syllables improves speech intelligibility almost as much as if the output noise level did not rise when speech is present. Indeed, for design purposes, it can be assumed that the apparent output noise level with speech present is only 5 db higher than the actual noise level with speech absent.[35]



a. Gain versus input power



b. Output power versus input power

Fig. 31—Characteristics of a typical expandor

This assumption may be used to relate the apparent SNR, K_{Ci}^{CE}, at the output of the i^{th} channel expandor to the actual SNR, K_{Ci}^{C}, of the compressed signal at the expandor input.* Let S_{Ci}^{CE} be the average speech power at the expandor output, and let N_{Ci}^{CE} be the average noise power when speech is absent. In terms of these symbols, the assumption becomes

$$K_{Ci}^{CE} = \frac{1}{3} S_{Ci}^{CE} / N_{Ci}^{CE}$$

If S_{Ci}^{C} and N_{Ci}^{C} are the signal and noise powers at the expandor input which correspond respectively to S_{Ci}^{CE} and N_{Ci}^{CE}, it follows that

$$10 \log \left(S_{Ci}^{C} / N_{Ci}^{C} \right) = \tfrac{1}{2} \, 10 \log \left(S_{Ci}^{CE} / N_{Ci}^{CE} \right)$$

that is, the difference in db between power levels at the expandor output is just twice that at the expandor input. But note that, while N_{Ci}^{C} was defined as the input noise when speech is absent, the input noise has the same value when speech is present. Therefore, S_{Ci}^{C} / N_{Ci}^{C} is just the SNR, K_{Ci}^{C}, at the expandor input. Making this substitution, and combining the last two equations, yields the desired relation between the SNRs at the input and output of an expandor

$$K_{Ci}^{C} = \sqrt{3K_{Ci}^{CE}} \qquad (271)$$

Note that this relation is independent of the crossover level of the expandor; it is true for all expandors with an expansion ratio of two, providing only that the input signal and noise levels both lie within the device's "range of levels."

*The superscript "C" indicates compression only, while "CE" implies compression followed by expansion.

Companding Improvement Factor

As with preemphasis, the cost-effectiveness of companding can be measured by comparing a link with companding to one without companding on the basis of equal single channel output SNRs. On the link without compandors, the average RF power which must be received to produce an output SNR, K_{Ci}, in the i^{th} channel is given by Eq. (249)

$$S_R = C\,\gamma_i\,K_{Ci} \tag{272}$$

where C is given by Eq. (225) and γ_i by Eq. (250). A similar relation holds for the companded link

$$S_R^C = C^C\,\gamma_i^C\,K_{Ci}^C \tag{273}$$

where the factors C^C and γ_i^C are given by expressions similar to Eqs. (225) and (250) but with superscripts "C" to indicate that they apply to the compressed signals and the parameters of the companded link. But note here that K_{Ci}^C is the SNR at the input to the i^{th} expandor, whereas it is the expandor output SNR K_{Ci}^{CE} that should be matched to the SNR objective K_i^o for the channel. Using Eq. (271) to express the former in terms of the latter, and taking the ratio of Eqs. (273) to (272), the improvement factor for companding is

$$I_C \equiv \left(\frac{S_R^C}{S_R}\right)_{K_{Ci}^{CE} = K_{Ci} = K_i^o} = \sqrt{3}\,\frac{C^C}{C}\,\frac{\gamma_i^C}{\gamma_i}\,\frac{1}{\sqrt{K_i^o}} \tag{274}$$

Thus, the companding improvement depends through C^C/C on the modulation method, modulation index, and baseband peak-to-average ratio, and through γ_i^C/γ_i on the baseband load factors. In contrast to the situation with preemphasis, it also depends on the signal-to-noise objective K_i^o

specified for the channel. Generally speaking, the higher the SNR objective the greater the value of companding. Note also that in applying Eq. (274) to angle modulation methods, the conditions on deviation ratio given by Eqs. (243-246) must be considered.

TIME ASSIGNMENT SPEECH INTERPOLATION (TASI)

TASI is a system developed by the Bell Telephone Laboratories for use on submarine telephone cables.* First installed in 1960, it has effectively doubled the capacity of those cables to which it has been connected by using the normal gaps in speech to interpolate additional conversations. There appear to be no serious technical barriers to the use of TASI on communication satellite links. In considering such use, however, it must be recognized that TASI offers very little gain in capacity unless the number of voice channels to be transmitted over a single link is greater than about twenty. Moreover, the high cost of the required terminal equipment makes TASI economically attractive only where the cost of increasing capacity in other ways is even higher, or where the spectrum required by alternative approaches is unavailable.

The operation of a communication link employing TASI will be explained with the aid of the block diagram in Fig. 32. Here, Q incoming telephone trunks are fed into the transmitting TASI unit which selectively connects them to a smaller number Q_T of channels at its output. These Q_T channels then form the inputs to a multichannel transmission link which may be a submarine cable or a radio link of the type shown in Fig. 1.

*The ensuing description is based largely on information from Refs. 37, 38, and 39.

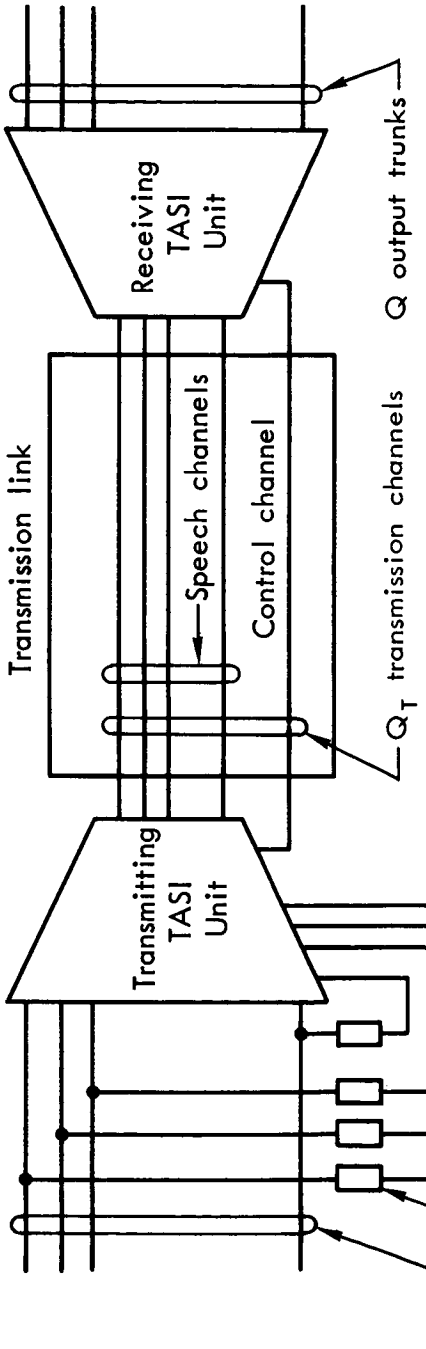


Fig. 32—TASI equipment for one direction of transmission

Whatever its nature, the link serves to multiplex and transmit the Q_T channels to the receiving terminal where, after demultiplexing, they are fed into the receiving TASI unit. This unit switches the signals arriving on its Q_T input channels among the Q telephone trunks at its output in such a way that each output trunk carries the waveform which arrived on its counterpart trunk at the transmitting TASI input.

That Q_T voice-bandwidth channels in the communication link can carry the conversations arriving on a larger number Q of telephone trunks is due to the fact that each of the input trunks is active (actually carrying speech power) only a fraction of the time. As explained in Section III, the average value of this fraction measured on a large number of channels during the busiest traffic hour is called the activity factor and has a value of about 0.25 on domestic trunks.[23] For overseas trunks, the activity factor is apparently somewhat higher and ranges from 0.35 to 0.40.

To detect activity, each of the Q incoming trunks is equipped with a speech detector. When the signal in a given channel exceeds a threshold of -40 dbm for at least 5 msec (to avoid triggering by noise spikes), the detector initiates a request for one of the transmission channels in the link. If an idle channel is available, the transmitting TASI unit assigns it to the requesting trunk. Before the actual connection is made, a "connect" signal is sent over the assigned channel to tell the receiving TASI unit which output trunk should be used. During the 17 msec required to connect talker and listener, the talker's speech is suppressed or "clipped" but the ear is relatively insensitive to such a short omission.

Once a transmission channel has been assigned to an incoming trunk, the connection is maintained until the speech power in the trunk drops below the speech detector threshold for more than 240 msec, known as the hangover time. Even then, in order to reduce the amount of switching during periods of light traffic, the connection is not broken unless the channel is needed for another trunk. When the channel is needed after a pause which exceeds the hangover time it is disconnected and a disconnect signal is sent to the receiving TASI unit over one of the Q_T transmission channels reserved for control purposes (see Fig. 32).

When speech is resumed in the trunk, a new channel request cycle is initiated. If, as occasionally happens, an idle transmission channel is not immediately available, speech on the incoming trunk is lost until a channel can be assigned. This form of speech clipping is known as "freezeout" and constitutes the most serious impairment to output signal quality caused by TASI. The total fraction of speech lost as a result of freezeout is called the "freezeout fraction." A freezeout fraction of two percent is objectionable to nearly all subscribers while a value of 0.1 percent is not noticeable. A freezeout fraction of 0.5 percent is considered satisfactory for design purposes.

The number of transmission channels Q_T (including one for control purposes) required to service Q telephone trunks depends on the activity factor τ and the freezeout factor. With a freezeout factor of 0.5 percent, values of Q_T versus Q are plotted in Fig. 33 for $\tau = 0.25$, 0.35, and 0.50. The ratio Q/Q_T is known as the TASI advantage and for $\tau = 0.35$ is seen to increase from about 1.7 for $Q_T = 10$ to 2.6 for $Q_T = 100$. Note that the TASI advantage approaches unity as the number of available transmission channels approaches one, whereas for large values of Q_T, the

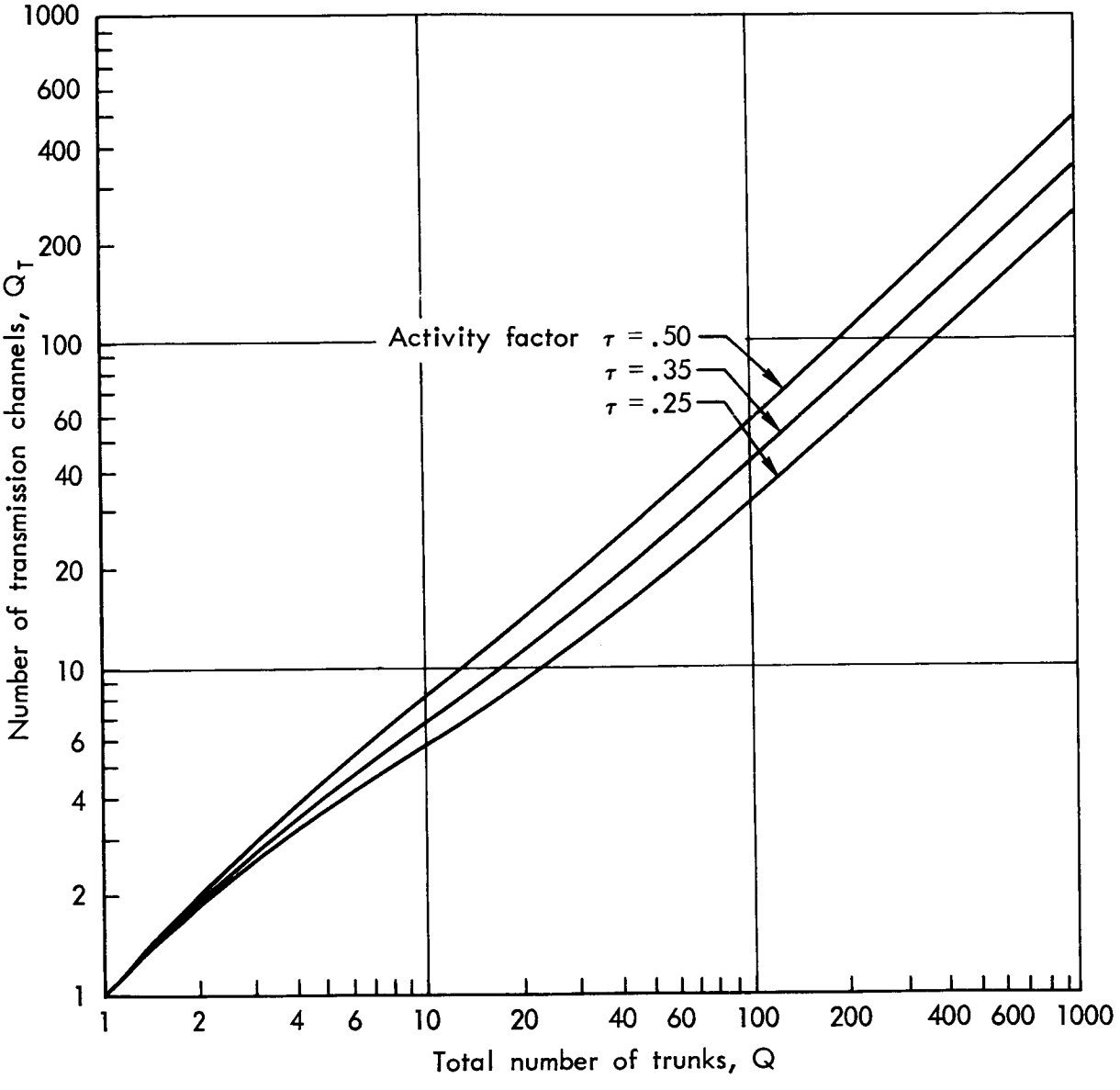


Fig. 33—Number of TASI transmission channels versus number of trunks

TASI advantage approaches $1/\tau$ asymptotically.

In the submarine cable application for which TASI was originally designed, the objective was to double the number of telephone trunks that could be handled by an expensive existing transmission link of fixed capacity. TASI could be used in a similar way on existing satellite links, but in considering such use it should be recognized that, in contrast to submarine cables, satellite links will for some time probably use modulation methods for which the transmission bandwidth is considerably greater than the baseband or information bandwidth. Thus, as an alternative to TASI on such links, more telephone trunks can be accomodated within the same RF bandwidth by merely reducing the modulation index. This approach will of course lead to some reduction in output signal quality since the same satellite transmitter power is now shared by more channels. As shown by Table 7, the output SNR, K_{Ci}, in the i^{th} channel, is proportional directly to the square of the modulation index and inversely to the product of baseband peak-to-average power ratio, Λ_B, and baseband load factor, $\Gamma_i = S_B/S_{Bi}$. For example, to effect a two-fold increase in baseband bandwidth with no change in RF bandwidth on an FM link, Carson's rule, Eq. (125), shows that the modulation index must be reduced from D to $\frac{1}{2}$ (D - 1). If the original number of channels is in the order of 30, Fig. 19 shows that doubling the number of channels will increase Γ_i by about 1.5 db, while Fig. 21 shows that Λ_B will be reduced by about 0.7 db. If the original modulation index was high, the overall reduction in output SNR will thus be about 6 + 1.5 - 0.7 = 6.8 db.

If TASI were used to obtain the same two-fold increase in the number of telephone trunks served by the satellite link, there would

also be a deterioration in output signal quality but it would not be as severe. In this case the baseband bandwidth and the modulation index would remain the same, but the activity factor, τ, would be increased from about 0.4 to 0.85. As a result, it is estimated that the load factor would be increased by about 2 db, while the peak-to-average ratio would again decrease by about 0.7 db. Thus, the output signal-to-noise ratio would decrease by about 2 - 0.7 = 1.3 db, or about 5.5 db less than the decrease resulting from changing the modulation index. Note that this comparison applies to links without preemphasis or companding; when one or both of these techniques are used the comparison can be expected to be slightly different. Also note that only the effect of TASI on output signal-to-noise ratio is considered, without regard to its effect on other aspects of signal quality.

Although TASI was originally designed to increase the effective capacity of existing links with fixed transmission bandwidths, it need not be restricted to this application. It can also be considered in planning new satellite systems as a means for reducing the required RF power and/or RF bandwidth when compared with a conventional link designed to carry the same number of telephone trunks and yield the same output signal quality. In this case, TASI leads to a reduction in baseband bandwidth at the price of an increase in activity factor. Neither the load factor, Γ_i, nor the peak-to-average power ratio, Λ_B, are changed significantly. As a result, if the same modulation index is used, the RF power will remain the same and the RF bandwidth will be reduced in the same proportion as the baseband bandwidth.

Alternatively, if one of the angle modulation methods is used, some or all of the RF bandwidth savings can be traded in the usual way for a

reduction in RF power. Thus, if TASI reduces the required baseband bandwidth by half, the modulation index can be more than doubled for a given RF bandwidth, provided that this increase does not bring the RF signal-to-noise ratio below its threshold value. Since the required RF power for a given output SNR is proportional to the inverse square of the modulation index, TASI can thus lead to about a 6 db reduction in RF power with no change in RF bandwidth.

Appendix A

COMPLEX WAVEFORMS AND FOURIER SPECTRA

A complex waveform is simply a complex-valued function of time

$$z(t) = x(t) + j\ y(t) \tag{A-1}$$

whose real and imaginary parts are ordinary real waveforms of the type used to represent time varying scalar physical quantities such as signal voltages or currents. A useful geometric picture of the complex waveform $z(t)$ is obtained by visualizing it as a two-dimensional vector whose rectangular components are $x(t)$ and $y(t)$. Such a vector may also be expressed in terms of its polar coordinates $r(t)$, $\theta(t)$

$$z(t) = r(t)\ e^{j\theta(t)}$$

The length $r(t)$ of the vector is given by the modulus of the complex waveform

$$r(t) = |z(t)| = \sqrt{x^2(t) + y^2(t)}$$

and the polar angle $\theta(t)$, measured counterclockwise from the real or x-axis, is given by the argument of $z(t)$

$$\theta(t) = \arg z(t) = \tan^{-1}[y(t)/x(t)]$$

From this point of view, the complex waveform is in general a vector of both variable length and variable angle.

The introduction of complex waveforms is motivated by the fact that the mathematical analysis of a purely real waveform can usually be simplified by substituting for it a properly chosen complex waveform whose

real part* is the real waveform in question. The complex waveform is then said to represent the real waveform--for example, in Eq. (A-1), z(t) represents x(t). The representation is ambiguous, however, until the time-dependence of the imaginary part of the complex waveform is specified. For example, in Eq. (A-1), y(t) must be specified in order to make z(t) a unique representation of x(t). While in principle the selection of y(t) is arbitrary, the desired mathematical simplicity is normally realized only for one or two particular choices.

The most common example of the use of complex waveforms arises in the representation of simple sinusoids of the form

$$x_o(t) = A \cos (\omega t + \varphi) \tag{A-2}$$

where A, ω, and φ are constants. If $x_o(t)$ is the real part of the complex waveform representation, then an obvious choice for the imaginary part is

$$y_o(t) = A \sin (\omega t + \varphi)$$

This makes the complex waveform an exponential function of time

$$z_o(t) = x_o(t) + jy_o(t) = Ae^{j(\omega t + \varphi)} \tag{A-3}$$

which is not only simple to manipulate mathematically, but has a very simple geometrical interpretation as a vector of constant length

$$|z_o(t)| = A$$

that rotates in a counterclockwise direction with uniform angular

*The imaginary part could equally well be used but the more common practice of using the real part will be followed here.

velocity ω. At any time t, the angle between this representative vector and the x-axis is

$$\arg z_o(t) = \omega t + \varphi$$

which is recognized as the phase angle of the original sinusoid.* The angular velocity of the vector gives the frequency $f = \omega/2\pi$ of the sinusoid, and the initial value of the vector, i.e., the value of $z_o(t)$ at $t = 0$, is a constant vector $Ae^{j\varphi}$ whose length and polar angle respectively give the peak amplitude A and initial phase φ of the sinusoid.

An equally satisfactory complex representation of the sinusoid $x_o(t)$ is obtained by setting the imaginary part equal to $-y_o(t)$. The complex waveform is then the complex conjugate of $z_o(t)$

$$z_o^*(t) = x_o(t) - j\, y_o(t) = Ae^{-j(\omega t + \varphi)} \qquad \text{(A-4)}$$

Note that $z_o^*(t)$ is also a rotating vector. It has the same constant length and speed of rotation as $z_o(t)$, but its initial angle is $-\varphi$ rather than φ, and its direction of rotation is opposite to that of $z_o(t)$--i.e., its angular velocity is negative, or clockwise. At all times, the vector $z_o^*(t)$ is the mirror image in the x-axis of $z_o(t)$.

Finally, it is obvious from Eqs. (A-3) and (A-4) that the real sinusoid $x_o(t)$ may be represented not only as the x-components of the rotating vectors $z_o(t)$ and $z_o^*(t)$, but also as the vector sum

$$x_o(t) = \tfrac{1}{2}\left[z_o(t) + z_o^*(t)\right] \qquad \text{(A-5)}$$

The utility of this representation of a sinusoid in terms of counter-rotating vectors (or conjugate complex waveforms) is illustrated in

---

*For this reason, the rotating vector is sometimes called a phase vector or "phasor."

the discussion of Fourier spectra that follows.

The concept of a Fourier spectrum is based on the fact that under certain rather general conditions, which are usually satisfied by physical signals, an arbitrary real waveform x(t) can be written as the sum (or integral) of an infinite number of simple sinusoids with properly chosen frequencies and amplitudes. The most familiar example arises when x(t) is periodic with period T. In this case, the Fourier sum is the infinite series

$$x(t) = x(t + T) = A_o + 2\sum_{n=1}^{\infty} (A_n \cos \omega_n t + B_n \sin \omega_n t) \qquad \text{(A-6)}$$

where the angular frequencies of the component sinusoids are

$$\omega_n = 2\pi n/T \qquad n = 1,2,3,\ldots$$

and their amplitudes, the so-called Fourier coefficients, may be calculated directly from x(t) by means of the formulas

$$A_n = \frac{1}{T}\int_{-T/2}^{T/2} x(t) \cos \omega_n t \, dt, \qquad B_n = \frac{1}{T}\int_{-T/2}^{T/2} x(t) \sin \omega_n t \, dt \qquad \text{(A-7)}$$

The coefficients A_n and B_n are thus the averages over the period T of the products of x(t) with cos $\omega_n t$ and sin $\omega_n t$ respectively. In a sense, they tell how much of the frequency ω_n is contained in x(t). By simple trigonometric manipulation, the series may also be written

$$x(t) = C_o + 2\sum_{n=1}^{\infty} C_n \cos (\omega_n t + \varphi_n) \qquad \text{(A-8)}$$

where the amplitudes C_n and phases φ_n can be calculated from the Fourier coefficients just given

$$C_n \equiv \sqrt{A_n^2 + B_n^2} \qquad \varphi_n \equiv \tan^{-1}(B_n/A_n) \qquad \text{(A-9)}$$

In either case, the set of component sinusoids is called the Fourier spectrum of x(t). The spectrum and hence x(t) itself is completely specified by giving the A_n and B_n or the C_n and φ_n as functions of angular frequency ω_n. Since only discrete frequencies are involved, the spectrum of a periodic function is called a line spectrum. Note that all of the component sinusoids are real time functions and their frequencies $\omega_n/2\pi$ are all positive numbers--positive integer multiples (harmonics) of a fundamental frequency $f_1 = T^{-1}$ defined by the period of x(t).

If each of the component sinusoids in the Fourier series of Eq. (A-8) is now replaced by its representation in terms of counterrotating vectors as in Eq. (A-5), straightforward algebraic manipulation yields

$$x(t) = \sum_{n=-\infty}^{\infty} X_n \, e^{j\omega_n t} \tag{A-10}$$

where again $\omega_n = 2\pi n/T$, and the X_n are complex numbers, calculable from x(t) by the formula

$$X_n = \frac{1}{T} \int_{-T/2}^{T/2} x(t) \, e^{-j\omega_n t} dt \tag{A-11}$$

The vector interpretation is quite straightforward. The Fourier series for the real waveform x(t) is now viewed as a sum of rotating vectors $X_n e^{j\omega_n t}$ where $n = 0, \pm 1, \pm 2, \ldots$. For any given value of t, these vectors may be summed using the parallelogram rule and the resultant will be a vector of length x(t) parallel to the real or x-axis. The n^{th} vector in the sum has length $|X_n|$ and rotates with a uniform angular velocity ω_n; at t = 0, it makes an angle with the real axis given by arg X_n. Thus the vector spectrum and hence x(t) itself is completely specified by giving the set of fixed initial vectors X_n as a function of angular velocity $\omega_n = 2\pi n/T$.

Since n ranges from $-\infty$ to $+\infty$, it is apparent that both positive (counterclockwise) and negative (clockwise) angular velocities are involved. This is only to be expected of course since the complex or vector Fourier series was obtained by replacing each sinusoid in its trigonometric counterpart by a pair of counterrotating vectors. By virtue of the close relation between the angular velocity ω_n of a rotating vector and the frequency $|\omega_n/2\pi|$ of the sinusoid that it represents, the vector angular velocities are often referred to as angular or radian frequencies.* The term "negative frequency," which still gives rise to occasional confusion in the literature, is merely a shorthand way of referring to the angular velocity of the clockwise rotating components in the rotating vector spectrum of a waveform.

It should be obvious that the complex exponential or rotating vector spectrum of Eq. (A-10) conveys, in a different form, exactly the same information as the sinusoidal spectrum of Eq. (A-8). This fact is confirmed by the close relationship between the initial vectors of Eq. (A-11) and the Fourier coefficients of Eqs. (A-7) and (A-9):

$$X_n = A_n + jB_n = C_n e^{j\varphi_n} \qquad \text{(A-12)}$$

Thus, the spectrum of a real periodic function x(t) is specified by pairs of real numbers $|X_n| = C_n$ and $\arg X_n = \varphi_n$, and it makes little difference whether a given pair is regarded as the amplitude and initial phase of a real sinusoid or as the length and initial polar angle of a rotating vector. However, despite this equivalence, the complex or rotating vector representation is today used almost exclusively in waveform

*The same term is of course also used in connection with a real sinusoid of frequency f to denote the product $2\pi f$ even where no rotating vector representation is involved.

analysis because it is both notationally and computationally simpler than its trigonometric counterpart.

When the real waveform x(t) is <u>not</u> periodic it still can be represented within any finite interval (-T/2, T/2) by the Fourier series of Eq. (A-10) with coefficients calculated from x(t) by Eq. (A-11). However, since the series is periodic with period T, it cannot represent x(t) outside the time interval.

To represent a nonperiodic x(t) over all time, the interval length T is allowed to approach infinity. As this is done of course, the fundamental angular velocity $\omega_1 = 2\pi/T$, the difference $\Delta\omega = \omega_1$ between the angular velocity of successive vectors, and the lengths $|X_n|$ of any given vector in the Fourier series all approach zero. On the other hand, if x(t) has finite energy--if the integral $\int_{-\infty}^{\infty} x^2(t)dt$ exists*--then the product TX_n given by Eq. (A-11) will remain finite as $T \to \infty$. As a result, the Fourier series (rewritten slightly by multiplying each term by $T\Delta\omega/2\pi = 1$)

$$x(t) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \left(TX_n\right) e^{j\omega_n t} \Delta\omega$$

becomes in the limit as $T \to \infty$ and $\Delta\omega = \omega_1 = 2\pi/T \to 0$, the Fourier integral

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega)\, e^{j\omega t} d\omega \qquad \text{(A-13)}$$

The function $X(\omega) \equiv \lim_{T\to\infty} TX_n$ is called the Fourier transform of x(t) and, by Eq. (11), is given by**

$$X(\omega) = \int_{-\infty}^{\infty} x(t)\, e^{-j\omega t} dt \qquad \text{(A-14)}$$

---

*This is a sufficient, but not a necessary condition for the existence of the Fourier spectrum.

**The Fourier transform of a waveform will usually be denoted by the upper case form of the symbol used for the waveform. When even more explicit identification of the waveform is desired, the operator notation $\mathfrak{F}$ will be used--e.g., $X(\omega) = \mathfrak{F}[x(t)]$.

From the vector point of view, the Fourier integral is just the sum of a continuous spectrum of infinitesimal rotating vectors, whose initial values (length and direction at t = 0) are determined by the Fourier transform $X(\omega)$. Like X_n, $X(\omega)$ affords an equivalent but computationally simpler way of describing the purely real trigonometric spectrum of a real time function.

The concept of a Fourier spectrum is by no means limited to real waveforms, however; it is easily extended to complex or vector waveforms

$$z(t) = x(t) + jy(t)$$

For example, if the real waveforms $x(t)$ and $y(t)$ are periodic with period T and have Fourier coefficients X_n and Y_n, then replacement of $x(t)$ and $y(t)$ by their Fourier series, yields

$$z(t) = \sum_{n=-\infty}^{\infty} X_n e^{j\omega_n t} + j \sum_{n=-\infty}^{\infty} Y_n e^{j\omega_n t} \equiv \sum_{n=-\infty}^{\infty} Z_n e^{j\omega_n t} \qquad \text{(A-15)}$$

where

$$Z_n \equiv X_n + j\, Y_n = \frac{1}{T} \int_{-T/2}^{T/2} z(t)\, e^{-j\omega_n t}\, dt \qquad \text{(A-16)}$$

is called the Fourier coefficient of $z(t)$. In exactly the same way, it follows that, if $z(t)$ is not periodic but has finite total energy $\int_{-\infty}^{\infty} |z(t)|^2 dt$, then it can be written as a Fourier integral

$$z(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Z(\omega)\, e^{j\omega t} d\omega \qquad \text{(A-17)}$$

where

$$Z(\omega) = \int_{-\infty}^{\infty} z(t)\, e^{-j\omega t} dt \qquad \text{(A-18)}$$

is the Fourier transform of z(t).

The fact that the Fourier relations for a complex waveform are formally identical to those for a real waveform constitutes another advantage of the complex Fourier representation over the trigonometric form. But where the use of the complex Fourier representation was a mathematical convenience in the case of real waveforms, it is a necessity for complex waveforms. The vector interpretation of the Fourier relations shows that a complex or vector function of time z(t) can be represented by an infinite sum or integral of rotating vectors whose initial lengths and angles are given as functions of their angular velocity by the Fourier coefficients Z_n or transform $Z(\omega)$ depending on whether x(t) is periodic or not. Since the constant vectors Z_n or $Z(\omega)$ define the spectrum of rotating vectors which in turn defines the waveform itself, they will in the future be referred to as simply "the spectrum" of the waveform.

As a complex (vector) function of angular velocity ω only, the spectrum of any real or complex waveform may be described graphically by separately plotting its modulus (length) and argument (polar angle) versus ω. When the waveform is real, these plots provide an easily interpreted description of the frequency composition of the waveform. As previously noted, the modulus and argument of the spectrum, evaluated for angular velocity ω, respectively give the amplitude and initial phase of the sinusoid with frequency $|\omega/2\pi|$ in the real trigonometric Fourier spectrum of the waveform. For this reason, the modulus plot is referred to as the amplitude spectrum, and the argument plot as the phase spectrum. Also, in accordance with popular usage, the vector

angular velocity will henceforth be called angular frequency or simply frequency.*

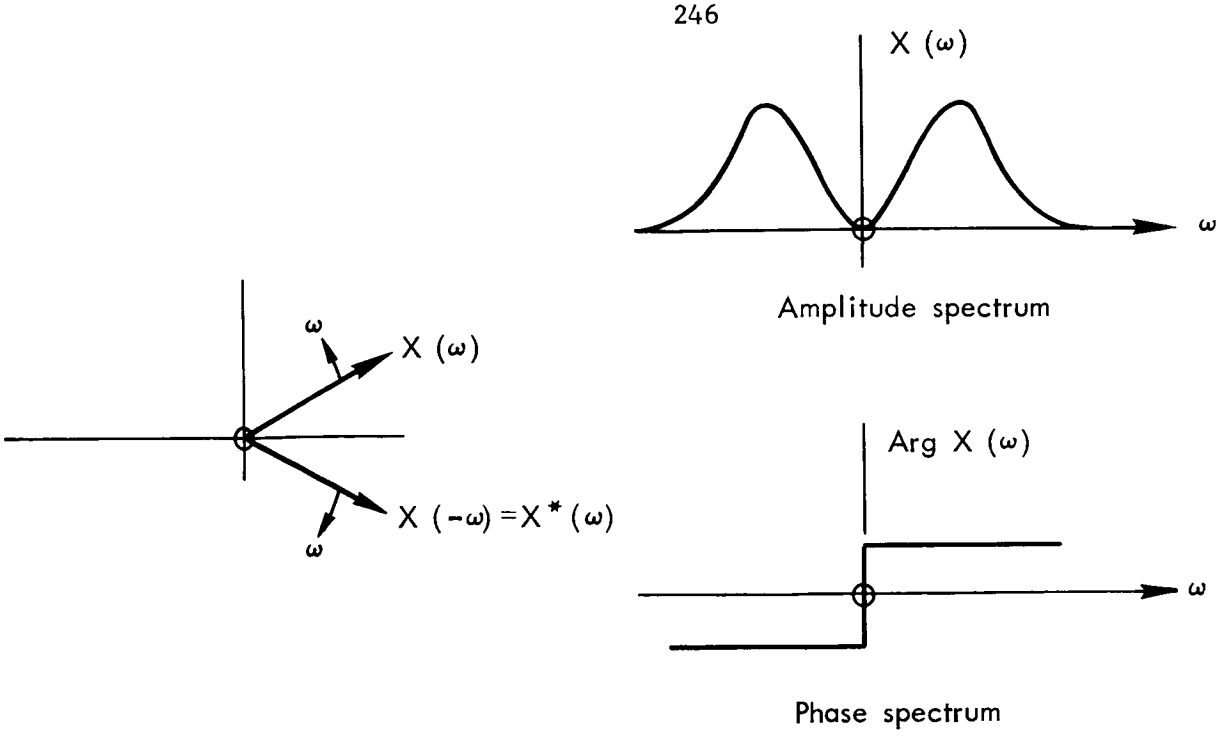
A striking feature of the spectral plots of all real waveforms is their symmetry about zero frequency. Thus, as illustrated in Fig. A1a, the amplitude spectrum for negative frequencies is a mirror image of the positive frequency amplitude spectrum reflected in the $\omega = 0$ axis. Likewise, the phase spectrum for negative frequencies is just the negative mirror image of the positive frequency phase spectrum. This symmetry is another consequence of the fact that each sinusoid in the trigonometric spectrum is represented in the (complex) spectrum by the sum of a pair of counterrotating vectors of equal length, equal angular speed, and equal but opposite initial phases; the vector rotating with frequency ω is just the reflection in the real axis of the vector rotating with frequency $-\omega$. Since the spectrum is nothing more than the set of these vector pairs evaluated at time zero, it is not surprising that in the special case of a real waveform, Eqs. (A-16) and (A-18) for the spectrum of an arbitrary complex waveform yield the formal description of this mirror symmetry:

$$Z_{-n} = Z_n^*, \text{ or } Z(-\omega) = Z^*(\omega) \qquad z(t) \text{ real}$$

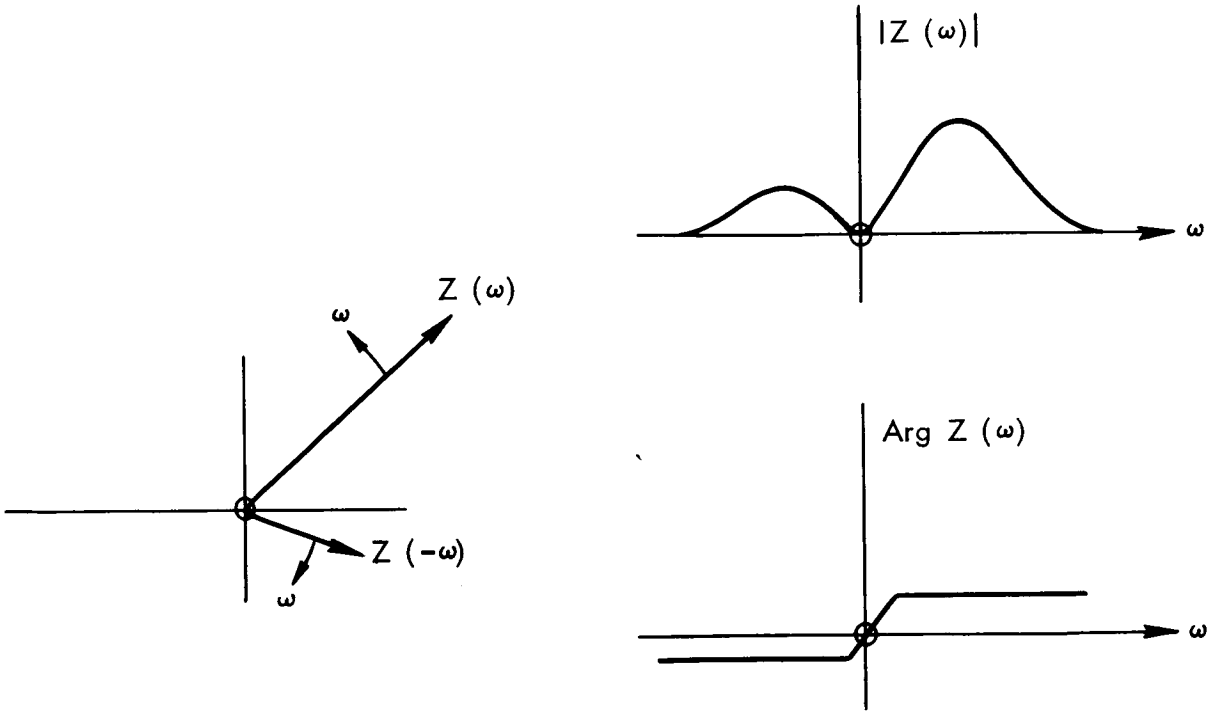
An example of this property of the spectra of real waveforms is also shown in Fig. A1a.

When the waveform is complex, spectral symmetry is necessarily lost; otherwise the rotating vectors would add up to a resultant parallel with the real axis--a real waveform. Instead, either the amplitude

*Thus, as noted earlier, "negative frequency" will mean clockwise angular velocity.



a. Real waveforms



b. Complex waveforms

Fig. A-1—Symmetry properties of Fourier spectra

or the phase spectra (or both) become unsymmetrical about $\omega = 0$, as illustrated for an arbitrary complex waveform in Fig. A1b. Indeed it is just this lack of spectral symmetry that makes complex waveforms especially useful in modulation theory.

The nature of the spectral asymmetry of a complex waveform depends of course on the relation between the spectra of its real and imaginary parts. Through proper choice of its imaginary part, the spectrum of the complex representation of a real waveform can be made to have almost any desired asymmetry. As a very useful example, if

$$z(t) = x(t) + jy(t)$$

is the complex representation of a nonperiodic real waveform $x(t)$ whose spectrum is $X(\omega)$, then the negative frequency spectrum of $z(t)$ can be made to vanish altogether. Since at all frequencies

$$Z(\omega) = X(\omega) + jY(\omega)$$

it is only necessary to choose for $y(t)$ a waveform whose negative frequency spectrum lags that of $x(t)$ by 90°

$$Y(\omega) = j\,X(\omega) \qquad \omega < 0 \qquad \text{(A-19)}$$

This not only makes $Z(\omega) = 0$ for $\omega < 0$ but, due to the mirror symmetry of the spectra of real waveforms, the positive frequency spectrum of $y(t)$ also exhibits a 90° phase lag:

$$Y(\omega) = -j\,X(\omega) \qquad \omega > 0 \qquad \text{(A-20)}$$

Thus the spectrum of $z(t)$ is not only one-sided but is equal to that of $2\,x(t)$ for positive frequencies, $Z(\omega) = 2X(\omega)$ for $\omega > 0$.

The special choice of $y(t)$ which produces these results is just $\breve{x}(t)$, the Hilbert transform of $x(t)$ introduced in Eq. (7). Its spectral properties, given in Eqs. (A-19) and (A-20), may be expressed more compactly in the form

$$\mathcal{F}\,[\breve{x}(t)] = -j\ \mathrm{sgn}\ (\omega)\ X(\omega) \tag{A-21}$$

where

$$\mathrm{sgn}(\omega) = \begin{cases} -1 & \omega < 0 \\ 0 & \omega = 0 \\ 1 & \omega > 0 \end{cases}$$

These properties provide another interpretation for the Hilbert transform of $x(t)$ as the response to $x(t)$ of a linear filter which introduces a 90° phase lag in all spectral components.

The corresponding complex waveform is called the analytic signal representation* of $x(t)$ and in this Memorandum will be denoted $\underline{x}(t)$:

$$\underline{x}(t) \equiv x(t) + j\ \breve{x}(t) \tag{A-22}$$

The spectrum of $\underline{x}(t)$ will be denoted $\underline{X}(\omega)$ and may be written

$$\underline{X}(\omega) = \begin{cases} 0 & \omega < 0 \\ X(\omega) & \omega = 0 \\ 2X(\omega) & \omega > 0 \end{cases}$$

or more compactly in several equivalent ways:

$$\underline{X}(\omega) = [1-\mathrm{sgn}(\omega)]\ X(\omega) = U(\omega)\ 2X(\omega) = 2X_{+}(\omega) \tag{A-23}$$

where $X_{+}(\omega) \equiv U(\omega)\ X(\omega)$ is the positive frequency spectrum of $x(t)$, and where $U(\omega)$ is the unit step function

$$U(\omega) = \begin{cases} 0, & \omega < 0 \\ \tfrac{1}{2}, & \omega = 0 \\ 1, & \omega > 0 \end{cases}$$

---

*See footnote accompanying Eq. (13).

Substituting Eq. (A-23) into Eq. (A-17) shows that the Fourier integral representation of the analytic signal $\underline{x}(t)$ may be viewed as the contribution of only the counterclockwise-rotating vectors (positive frequencies) in the Fourier spectrum of x(t):

$$\underline{x}(t) = \frac{1}{\pi}\int_0^\infty X(\omega)\, e^{j\omega t} d\omega \equiv 2x_+(t) \tag{A-24}$$

Here

$$x_+(t) \equiv \frac{1}{2\pi}\int_0^\infty X(\omega)\, e^{j\omega t} d\omega \tag{A-25}$$

is the so-called "positive frequency content"[10] of the real waveform x(t) represented by $\underline{x}(t)$. The spectrum of $x_+(t)$ is of course just $X_+(\omega)$, the positive frequency spectrum of x(t) introduced in Eq. (A-23).

The complex conjugate of the analytic signal

$$\underline{x}^*(t) = x(t) - j\check{x}(t) \tag{A-26}$$

forms another useful complex representative of x(t). Like $\underline{x}(t)$, its spectrum is one-sided, but in this case, it is the positive frequency portion that vanishes while the negative frequency part is equal to the negative frequency spectrum of 2x(t)

$$\mathcal{F}[\underline{x}^*(t)] = \underline{X}^*(-\omega) = \begin{cases} 2X(\omega) & \omega < 0 \\ X(\omega) & \omega = 0 \\ 0 & \omega > 0 \end{cases}$$

or more compactly

$$\mathcal{F}[\underline{x}^*(t)] = (1 - \operatorname{sgn}\omega)\, X(\omega) = U(-\omega)2X(\omega) = 2X_-(\omega) \tag{A-27}$$

where $X_-(\omega) = U(-\omega)X(\omega)$ is the negative frequency spectrum of x(t). Thus $\underline{x}^*(t)$ may be viewed as the contribution of only the clockwise-rotating vectors in the Fourier spectrum of x(t)

$$\underline{x}^*(t) = \frac{1}{\pi}\int_{-\infty}^{0} X(\omega)\, e^{j\omega t} d\omega \equiv 2x_-(t) \qquad \text{(A-28)}$$

where

$$x_-(t) = \frac{1}{2\pi}\int_{-\infty}^{0} X(\omega)\, e^{j\omega t} d\omega \qquad \text{(A-29)}$$

is called the negative frequency content[10] of $x(t)$, and has the spectrum $X_-(\omega)$.

It is useful to note that the rotating vector representation of a simple sinusoid introduced in Eq. (A-3) is in fact just its analytic signal representation. Indeed, it is now apparent that the analytic signal representation $\underline{x}(t)$ of an arbitrary waveform $x(t)$ may be viewed as the time-varying counterclockwise-rotating vector that results from replacing each sinusoid in the real trigometric form of the Fourier integral representation of $x(t)$ by its own analytic signal representation, i.e., a counterclockwise-rotating vector. Thus if in the Fourier integral analog to Eq. (A-8)

$$x(t) = \frac{1}{\pi}\int_{0}^{\infty} C(\omega)\cos\left[\omega t + \varphi(\omega)\right] d\omega \qquad \text{(A-30)}$$

the integrand is replaced with its analytic signal representation

$$C(\omega)\exp\left[\omega t + \varphi(\omega)\right] \equiv X(\omega)e^{j\omega t}$$

where

$$X(\omega) = C(\omega)e^{j\varphi(\omega)}$$

then the result is just the expression for the analytic signal of $x(t)$ given in Eq. (A-24). Similarly, replacing each sinusoid in Eq. (A-30) with the complex conjugate of its analytic signal leads to the expression for the complex conjugate of the analytic signal representation of $x(t)$ given in Eq. (A-28). In general $\underline{x}^*(t)$ is thus a clockwise-rotating vector whose length and angular velocity both vary with time.

Just as a simple sinusoid A cos $(\omega t + \varphi)$ can be expressed in terms of rotating vectors in three equivalent ways:

$$A\cos(\omega t + \varphi) \equiv x_o(t) = \begin{cases} \text{Re } \underline{x}_o(t) = \text{Re } Ae^{\,j(\omega t + \varphi)} \\ \text{Re } \underline{x}_o^*(t) = \text{Re } Ae^{\,-j(\omega t + \varphi)} \\ \frac{1}{2}\left[\underline{x}_o(t) + \underline{x}_o^*(t)\right] = \frac{A}{2}\left[e^{j(\omega t + \varphi)} + e^{-j(\omega t + \varphi)}\right] \end{cases}$$

so also can an arbitrary real time function be expressed in three ways in terms of the time-varying rotating vectors defined by its analytic signal $\underline{x}(t)$:

$$x(t) = \begin{cases} \text{Re } \underline{x}(t) \\ \text{Re } \underline{x}^*(t) \\ \frac{1}{2}\left[\underline{x}(t) + \underline{x}^*(t)\right] \end{cases}$$

It should be emphasized that not every complex waveform is the analytic signal representation of its real part. For example, if a(t) and $\varphi(t)$ are independent and arbitrary real waveforms, the function

$$c(t) = a(t)\exp\left[j(\omega_o t + \varphi(t))\right] \tag{A-31}$$

is a complex representation of the real modulated carrier

$$e(t) = a(t)\cos\left[\omega_o t + \varphi(t)\right] \tag{A-32}$$

but in general, it is not the analytic signal representation. The latter is given by

$$\underline{e}(t) = e(t) + j\check{e}(t) = a'(t)\exp\left[\omega_o t + \varphi'(t)\right] \tag{A-33}$$

where

$$\begin{aligned} a'(t) &= |\underline{e}(t)| = \sqrt{e^2(t) + \check{e}^2(t)} \\ \omega_o t + \varphi'(t) &= \arg \underline{e}(t) = \tan^{-1}\left[\check{e}(t)/e(t)\right] \end{aligned} \tag{A-34}$$

In order for c(t) to be the analytic signal it would be necessary for

$a(t) \sin [\omega_o t + \varphi(t)]$ to be the Hilbert transform of $a(t) \cos [\omega_o t + \varphi(t)]$, and this is not generally the case for arbitrary $a(t)$ and $\varphi(t)$.

On the other hand, it is true that any complex waveform whose spectrum vanishes for negative frequencies is the analytic signal representation of its real part. Likewise a complex waveform whose spectrum vanishes for positive frequencies is the conjugate of the analytic signal representation of its real part. Thus, in terms of the example in the last paragraph, if it can be shown that the spectrum of $c(t)$ vanishes for negative frequencies, it would follow that

$$c(t) = \underline{e}(t)$$

Analytic signals and Fourier transforms have many other interesting and useful properties but probably the most important of them for modulation theory is that the spectrum of the product of two real or complex waveforms is the convolution of their individual spectra. In symbols, if

$$z(t) = z_1(t)\, z_2(t)$$

then

$$Z(\omega) = Z_1(\omega) * Z_2(\omega)$$

where the convolution operation is defined as

$$Z_1(\omega) * Z_2(\omega) \equiv \int_{-\infty}^{\infty} Z_1(\Omega)\, Z_2(\omega - \Omega)\, d\Omega$$

In the special case

$$z_1(t) = m(t), \quad z_2(t) = e^{j\omega_o t}$$

$$z(t) = m(t)\, e^{j\omega_o t}$$

this yields

$$\mathfrak{F}[m(t)\, e^{j\omega_o t}] = Z(\omega) = M(\omega - \omega_o)$$

Thus, the effect of multiplying an arbitrary waveform, real or complex, by $e^{j\omega_o t}$ is to shift its spectrum upward in frequency by an amount ω_o. Similarly, the effect of multiplying m(t) by $e^{-j\omega_o t}$ is a downward frequency shift by ω_o

$$\mathfrak{F}\left[m(t)\ e^{-j\omega_o t}\right] = M(\omega + \omega_o)$$

It follows that if $M(\omega + \omega_o) = 0$ for $\omega < -\omega_o$ then z(t) is the analytic signal representation of

$$\mathrm{Re}z(t) = a(t) \cos\left[\omega_o t + \varphi(t)\right]$$

Finally, multiplication by

$$\cos \omega_o t = \tfrac{1}{2}\left[e^{j\omega_o t} + e^{-j\omega_o t}\right]$$

produces both an upward and a downward shift

$$\mathfrak{F}\left[m(t) \cos \omega_o t\right] = \tfrac{1}{2}\left[M(\omega-\omega_o) + M(\omega + \omega_o)\right]$$

REFERENCES

1. Bedrosian, E., N. Feldman, G. Northrop, and W. Sollfrey, Multiple Access Techniques for Communication Satellites: I. Survey of the Problem, The RAND Corporation, RM-4298-NASA, September 1964.

2. Lindholm, C. R., Multiple Access Techniques for Communication Satellites: Digital Modulation, Time-Division Multiplexing and Related Signal Processing, The RAND Corporation, RM-4997-NASA, September 1966.

3. Voelcker, Herbert, "On the Origin and Characteristics of Single-Sided Angle Modulation," Trans. IEEE, Vol. COM-13, No. 4, December 1965, pp. 555-556.

4. Bedrosian, E., "The Analytic Signal Representation of Modulated Waveforms," Proc. IRE, Vol. 50, No. 10, October 1962, pp. 2071-2076.

5. Titschmarsch, E. C., Introduction to the Theory of Fourier Integrals, Oxford University Press, New York, N. Y., 1937, Chap. 5.

6. Papoulis, Athanasios, Probability, Random Variables, and Stochastic Processes, McGraw-Hill Book Co., Inc., New York, 1965.

7. Booker, Henry G., A Vector Approach to Oscillations, Academic Press, New York, 1965.

8. Voelcker, Herbert B., "Toward a Unified Theory of Modulation Part I: Phase-Envelope Relationships," Proc. IEEE, Vol. 54, No. 3, March 1966, pp. 340-353.

9. Dugundji, J., "Envelopes and Pre-Envelopes of Real Waveforms," Trans. IRE, Vol. IT-4, No. 1, March 1958, pp. 53-57.

10. Rowe, H. E., Signals and Noise in Communication Systems, D. Van Nostrand Company Inc., Princeton, New Jersey, 1965.

11. Downing, John J., Modulation Systems and Noise, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1964.

12. Abramson, Norman, "Bandwidth and Spectra of Phase-and-Frequency-Modulated Waves," Trans. IEEE, Vol. CS-11, No. 4, December 1963, pp. 407-414.

13. Zakai, Moshe, "A Class of Definitions of 'Duration' (or 'Uncertainty') and the Associated Uncertainty Relations," Information and Control, Vol. 3, 1960, pp. 101-115.

14. Stewart, John L., "The Power Spectrum of a Carrier, Frequency Modulated by Gaussian Noise," Proc. IRE, Vol. 42, No. 10, October 1954, pp. 1539-1542.

15. Bedrosian, E., Distortion and Crosstalk of Linearly Filtered, Angle-Modulated Signals, The RAND Corporation, RM-4888-NASA, March 1966.

16. Voelcker, Herbert, "Demodulation of Single-Sideband Signals Via Envelope Detection," Trans. IEEE, Vol. COM-14, No. 1, February 1966, pp. 22-30.

17. Viterbi, A. J., "Phase-Lock-Loop Systems," in A. V. Balakrishman (ed.), Space Communications, McGraw-Hill Book Co., Inc., New York, 1963, pp. 123-142.

18. Jaffe, R., and E. Rechtin, "Design and Performance of Phase-Lock Circuits Capable of Near-Optimum Performance Over a Wide Range of Input Signal and Noise Levels," Trans. IRE, Vol. IT-1, No. 1, March 1955, pp. 66-76.

19. Stumpers, F. L. H. M., "Theory of Frequency-Modulation Noise," Proc. IRE, Vol. 36, No. 9, September 1948, pp. 1081-1092.

20. Chaffee, J. G., "The Application of Negative Feedback to Frequency-Modulation Systems," Bell Sys. Tech. J. Vol. 18, No. 3, July 1939, pp. 404-437.

21. Bedrosian, E., Power-Bandwidth Trade-offs for Feedback FM Systems: A Comparison With Pulse-Code-Modulation, The RAND Corporation, RM-2787-NASA, October 1961.

22. Enloe, L. H., "Decreasing the Threshold in FM by Frequency Feedback," Proc. IRE, Vol. 50, No. 1, January 1962, pp. 18-30.

23. Holbrook, B. D., and J. V. Dixon, "Load Rating Theory for Multi-Channel Amplifiers," Bell Sys. Tech. J., Vol. 18, No. 1, October 1939, pp. 624-644.

24. Cyr, M. J., and A. Thuswaldner, "Multichannel Load Calculation Using the Monte Carlo Method," Trans. IEEE, Vol. COM-14, No. 2, April 1966, pp. 177-181.

25. Dunn, H. K., and S. D. White, "Statistical Measurements on Conversational Speech," J. Acoustical Soc. Am., Vol. 2, January 1940, pp. 278-288.

26. Davenport, Wilbur B., "An Experimental Study of Speech-Wave Probability Distributions," J. Acoustical Soc. Am., Vol. 24, No. 4, July 1952, pp. 390-399.

27. Purton, R. F., A Survey of Telephone Speech-Signal Statistical and Their Significance in the Choice of a P.C.M. Companding Law, The Institution of Electrical Engineers, Paper No. 3773E, January 1962.

28. Subrizi, V., "A Speech Volume Survey on Telephone Message Circuits," Bell Labs. Rec., August 1953, pp. 292-295.

29. "C.C.I.T.T. Collected Documents on the Volume and Power of Speech Currents Transmitted Over International Telephone Circuits," IInd Plenary Assembly Red Book, Vol. III, International Telecommunication Union, 1962, pp. 343-350.

30. Brand, S., "The 'Vu' and the New Volume Indicator," Bell Labs. Rec., June 1940, pp. 310-314.

31. Smith, Herbert L., "The Use of Holbrook and Dixon Loading Factors in Setting Receiver Parameters for an FDM-FM Radio Telephone Communication System," Trans. IEEE, Vol. COM-12, No. 4, December 1964, pp. 155-161.

32. Oliver, W., "White Noise Loading of Multi-Channel Communication Systems," Electronic Engineering, November 1965, pp. 714-717.

33. Parry, Charles A., "CCITT Recommendations for Multichannel Radio Relay and White Noise," AIEE Trans. Commun. and Elec., Vol. 78, Pt. 1, No. 42, May 1959, pp. 107-117.

34. Carter, C. W., A. C. Dickieson, and D. Mitchell, "Application of Compandors to Telephone Circuits," Trans. AIEE, Vol. 65, Supplement for December 1946, pp. 1079-1086.

35. Rizzoni, Eitel M., "Compandor Loading and Noise Improvement in Frequency Division Multiplex Radio-Relay Systems," Proc. IRE, Vol. 48, No. 2, February 1960, pp. 208-220.

36. "Telephony Compandors," IInd Plenary Assembly Red Book, Vol. III, International Telecommunication Union, 1962, pp. 33-39.

37. Bullington, K., and J. M. Fraser, "Engineering Aspects of TASI," Bell Sys. Tech. J., Vol. 38, March 1959, pp. 353-364.

38. Fraser, J. M., D. B. Bullock, and N. G. Long, "Over-All Characteristics of a TASI System," Bell Sys. Tech. J., Vol. 41, July 1962, pp. 1439-1454.

39. Miedema, H. and M. G. Schachtman, "TASI Quality--Effect of Speech Detectors and Interpolation," Bell Sys. Tech. J., Vol. 41, July 1962, pp. 1455-1473.